Journal of Consulting and Clinical Psychology
1998, Vol. 66, No. 6, 914–923

Utility of Psychophysiological Measurement in the Diagnosis of Posttraumatic Stress Disorder: Results From a Department of Veterans Affairs Cooperative Study

Terence M. Keane
Veterans Affairs Medical Center, Boston, and
Boston University School of Medicine

Lawrence C. Kolb
Sea Island, Georgia

Danny G. Kaloupek
Veterans Affairs Medical Center, Boston, and
Boston University School of Medicine

Scott P. Orr
Veterans Affairs Medical Center, Manchester, and
Harvard Medical School

Edward B. Blanchard
University at Albany, State University of New York

Ronald G. Thomas
University of California, San Diego

Frank Y. Hsieh
Veterans Affairs Health Care System, Palo Alto

Philip W. Lavori
Veterans Affairs Health Care System, Palo Alto, and
Stanford University

This multisite study tested the ability of psychophysiological responding to predict posttraumatic stress disorder (PTSD) diagnosis (current, lifetime, or never) in a large sample of male Vietnam veterans. Predictor variables for a logistic regression equation were drawn from a challenge task involving scenes of combat. The equation was tested and cross-validated, demonstrating correct classification of approximately 2/3 of the current and never PTSD participants. Results replicate the finding of heightened psychophysiological responding to trauma-related cues by individuals with current PTSD, as well as differences in a variety of other domains between groups with and without the disorder. Follow-up analyses indicate that veterans with current PTSD who do not react physiologically to the challenge task manifest less reexperiencing symptoms, depression, and guilt. Discussion addresses the value of psychophysiological measures for assessment of PTSD.

The study of the psychophysiology of posttraumatic stress disorder (PTSD) can be traced to Kardiner (1941), who noted the elevated muscle tension, tachycardia, startle, and hyperresponsivity to stimulation that characterized war stress. This observation prompted empirical studies by Wenger (1948) and Dobbs and Wilson (1960), who demonstrated that physiological and psychophysiological measures could distinguish veterans with combat stress disorders from a variety of comparison groups. The contemporary foundation for this line of investigation can be traced to a pair of studies conducted in independent

Terence M. Keane and Danny G. Kaloupek, Psychology Service, Veterans Affairs Medical Center (VAMC), Boston, and Department of Psychiatry, Boston University School of Medicine; Lawrence C. Kolb, Sea Island, Georgia; Scott P. Orr, Research Service, VAMC, Manchester, and Department of Psychiatry, Harvard Medical School; Edward B. Blanchard, Center for Stress and Anxiety Disorders, University at Albany, State University of New York; Ronald G. Thomas, Department of Family and Preventive Medicine and Neuroscience, University of California, San Diego; Frank Y. Hsieh, Cooperative Studies Program Coordinating Center, VA Health Care System, Palo Alto; Philip W. Lavori, Cooperative Studies Program Coordinating Center, VA Health Care System, Palo Alto, and Department of Health Research and Policy, Stanford University.

This research was funded by the Department of VA Cooperative Studies Program of the Veterans Health Administration under Designation CS-334. Contributions were made by site investigators and members of the Cooperative Studies Program staff, as well as by members of the study Executive Committee and Data Monitoring Board (listed alphabetically): Francis Abueg, Aradanaha Bhat, Patrick Boudewyns, Dale Cannon, Diane Castillo, Dennis Charney, Claude Chemtob, Jonathan Davidson, Joseph Fleiss, Adele George, Robert Gerardi, Paul Ingmundson, Kenneth James, Chowdary Jampala, Michael Jueng, Stanislav Kasl, Nona Kay, Richard Lane, Judith Lyons, Barbara Melamed, Roger Pitman, Steven Price, Lenore Sheridan, Mark Slater, Edwin Smelker, Edward Snyder, Steven Southwick, Patricia Sutker, and Rose Zimering. Robert Rosenheck provided support to the project in the form of consultation and data entry services for the War Stress Interview.

Results of this study were presented at the meetings of the American Psychological Association, Los Angeles, California, August 1994; the European Conference on Traumatic Stress, Paris, France, May 1995; and the American Psychiatric Association, Miami, Florida, May 1995.

We thank the study clinicians, the technicians, and the veterans who participated in this study.

Correspondence concerning this article should be addressed to Terence M. Keane, National Center for Posttraumatic Stress Disorder, Boston VAMC (116B-2), Boston, Massachusetts 02130. Electronic mail may be sent to Keane.Terence@Boston.VA.Gov.

laboratories in the early 1980s. First, Blanchard, Kolb, Pallmeyer, and Gerardi (1982) found that psychophysiological responses to standardized sounds of combat discriminated Vietnam veterans with PTSD from an age- and gender-matched comparison group. Next, Malloy, Fairbank, and Keane (1983) used standardized audiovisual cues to demonstrate that male Vietnam veterans with PTSD were more responsive to the combat cues than either psychiatrically impaired veterans without PTSD or well-adjusted veterans. A few years later, Pitman, Orr, Forgue, de Jong, and Claiborn (1987) adapted idiographic methods (see Lang, Levin, Miller, & Kozak, 1983) for studying emotion to compare the psychophysiological responding of male Vietnam veterans with and without PTSD to individually tailored imagery scripts. The 30-s scripts contained stimulus, response, and meaning propositions that depicted, among others, the two most stressful combat experiences recalled by the veteran. Results demonstrated greater reactivity to imagery-based trauma cues for the PTSD group relative to the non-PTSD comparison group.

These and other studies involving psychophysiological challenge tasks have provided impressive preliminary evidence that PTSD patients can be discriminated from non-PTSD patients (see Orr & Kaloupek, 1997; Prins, Kaloupek, & Keane, 1995). Yet, there are methodological limitations in these studies that preclude definitive conclusions regarding diagnostic accuracy. A major limitation is the high PTSD base rate in studies that have used psychophysiological variables to classify participants relative to diagnostic status. Artificially high classification accuracy can occur when base rates of the targeted disorder are high relative to the population to which the measures will be later applied (see Kraemer, 1992; Tomarken, 1995). This is a concern with past studies that have enrolled PTSD participants as one third to one half of the total sample. A related problem is the presence of nonhelp-seeking comparison groups. Nearly all studies of reactivity to trauma-relevant cues include a major comparison group composed of well-adjusted combat veterans, typically volunteers. Comparison with these participants can inflate diagnostic hit rates by capitalizing on the myriad of differences irrespective of PTSD status. A third limitation of existing studies is the failure to adequately demonstrate cross-validation of the classification equations. With some notable exceptions (e.g., Blanchard, Kolb, & Prins, 1991; Orr, Pitman, Lasko, & Herz, 1993), most studies have not attempted cross-validation on an independent sample or had too few participants to adequately test stability of the prediction equations. Finally, the studies to date have not used multivariate analytic methods to investigate classification accuracy because sample sizes have been too small for these types of analyses.

For these reasons, especially the need for a large sample, a multisite clinical trial was initiated under the auspices of the Cooperative Studies Program of the Department of Veterans Affairs (DVA). The primary purpose of this study was to evaluate the extent to which psychophysiological measures can predict the presence or absence of PTSD as determined by a gold-standard diagnostic interview, the Structured Clinical Interview for *DSM–III–R* (SCID; Spitzer, Williams, Gibbon, & First, 1989). A sample size was projected to permit a complete utility analysis (i.e., sensitivity, specificity, predictive power of a positive test, and predictive power of a negative test) of the type required for all measures of psychiatric and medical disorders (Gerardi, Keane, & Penk, 1989; Kraemer, 1992).

Method

Participants

Participants were male military veterans who served in the Vietnam theater of operations between August 1964 and May 1975 and who were currently using services of DVA. Recruitment took place over a 42-month period from inpatient and outpatient programs in psychiatry, substance abuse, and PTSD, along with programs under Readjustment Counseling Service.

Individuals were excluded from participation if they were already involved in research sponsored by the Cooperative Studies Program or if they had physical conditions or used medications that might markedly alter their psychophysiological responding. Individuals remained eligible for participation if both they and their attending physician agreed to medication withdrawal and discontinuation for a period of four half-lives plus 14 days before completing the psychophysiological assessment procedures. Of the 2,115 individuals screened as potential participants, 1,461 qualified for eligibility.

Diagnostic and Psychometric Assessment

Interviews were conducted in fixed order by doctoral-level clinicians. The War Stress Inventory (Rosenheck & Fontana, 1989) was administered first to obtain sociodemographic, psychosocial history, and mental health information. This was followed by SCID interview modules for major depression, bipolar disorder, schizophrenia (psychotic screen), alcohol abuse and dependence, drug abuse and dependence, panic disorder, social phobia, obsessive–compulsive disorder, somatoform disorder, dissociative disorder, and PTSD (combat related and noncombat related). We assessed antisocial and borderline personality disorders using the SCID–II. All SCID interviews were audiotaped, and 128 were reviewed by a study clinician at a second site. Interrater reliability expressed as the percentage of agreement on PTSD diagnostic categories current, lifetime, or never was 77% (κ = 0.68, with weights of 0, 0.5, and 1.0). Similarly, a second clinician at the same site repeated PTSD interviews on 36 participants, resulting in a reliability score of 78% (κ = 0.66, with weights of 0, 0.5, and 1.0).

The interview was completed by 1,328 participants (778 current, 181 lifetime, and 369 never). Participants next completed the Minnesota Multiphasic Personality Inventory (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1990), the Keane PTSD (PK; Keane, Malloy, & Fairbank, 1984) scale, the Combat Exposure Scale (Keane et al., 1989), the Mississippi Scale for Combat-Related PTSD (Keane, Caddell, & Taylor, 1988), and the Laufer-Parsons Inventory (LPI; Laufer, Yager, Frey-Wouters, & Donnellan, 1981) to assess combat-related guilt.

The initial psychophysiological test session was completed by 1,210 participants. Forty-two records were eliminated during the processing of psychophysiological data, leaving a total of 1,168 participants available for analysis of psychophysiological variables (654 current, 154 lifetime, and 340 never PTSD).

Psychophysiological Assessment

Format. The psychophysiological challenge task included audiovisual presentation of standardized still images with accompanying soundtrack, modeled after the Malloy et al. (1983) study, and recorded scripts that depicted idiographic imagery scenes. Both neutral and combat audiovisual presentations consisted of six photographic images that had been recorded onto videotape. The neutral tape contained outdoor scenes that were intended to be distinct from Vietnam, with light classical piano music recorded on the audio track. The combat tape contained scenes from Vietnam depicting a helicopter assault and firefight and included sounds of helicopters, small arms fire, explosions, and voices of combatants. The volume level of each tape was set to approximately 65 dB at the participant's seated location.

Combat imagery scripts were composed by the participant with the assistance of the study clinician after all interviews and psychometric instruments were complete. A systematic method (see Pitman et al., 1987) was applied to generate individualized scripts portraying the two most stressful combat experiences from the participant's period of service in Vietnam. The two neutral scripts were standardized; one portrayed a quiet scene viewed from a lawn chair, the other described a scene at the beach.

Apparatus and measures. Physiological dependent measures included heart rate (HR), skin conductance (SC), left lateral frontalis electromyogram (EMG), and systolic and diastolic blood pressure (SBP and DBP, respectively). HR was recorded from 9-mm-diameter Sensor Medics Ag/AgCl electrodes filled with Beekman electrolyte paste and attached by adhesive collars at standard lead I (arm) sites. Electrodes were connected to a Coulbourn High Gain Bioamplifier (S75-01), and output from the amplifier was directed to a Coulbourn Tachometer (S77-26) to yield a beat-by-beat voltage that was proportional to interbeat interval. SC was measured directly by a Coulbourn Isolated Skin Conductance coupler (S71-23) using a constant 0.5-V output through 9-mm diameter Sensor Medics Ag/AgCl electrodes filled with an isotonic paste (Fowles et al., 1981). Electrodes were attached to the hypothenar surface of the nondominant hand, separated by 14 mm. EMG was recorded by use of 4-mm-diameter Sensor Medics Ag/AgCl surface electrodes filled with Beckman electrolyte paste and attached by adhesive collars over the left lateral frontalis muscle. Skin preparation and electrode placement followed published recommendations (Fridlund & Cacioppo, 1986). EMG electrodes were connected to a Coulbourn High Gain Bioamplifier (S75-01) set to filter signal components that were less than a 90-Hz or more than 250-Hz frequency. The EMG signal was directed to a Coulbourn Contour Following Integrator (S76-01) for integration at a setting of 300 ms. BP was measured using a Critikon Dinamap–automated BP monitor (Model 1846 SX) with the inflatable occlusive cuff (adult size, 23–33 cm) positioned on the dominant arm.

The HR, SC, and EMG analog signals were digitized by a Coulbourn Lablinc Analog-to-Digital Converter (L25-12), which was connected to an IBM-compatible computer through a Coulbourn Labline Computer Interface (L18-16). The Dinamap monitor for BP was connected to the computer through a serial port interface. Physiological signals were sampled at 2 Hz and converted to appropriate measurement units (i.e., beats per minute for HR, microsiemens for SC, and microvolts for EMG). The computer controlled the audiovisual presentations, the audio recorder, and the tone generator used for the imagery script procedure; generated rating scale displays on the video monitor; and recorded rating values registered through a joystick.

Procedures. BA-level technicians were responsible for administering the 90-min psychophysiological protocol that began with collection of a urine specimen. Participants were then escorted to the testing room and seated in a recliner while electrodes and the BP cuff were attached. Testing began with a 10-min rest period (Baseline 1) in which HR, SC, and EMG were recorded during the last 5 min. BP was sampled once at the immediate end of the period.

Next, participants performed mental arithmetic (the generic stressor) for 2 min. HR, SC, and EMG were recorded throughout the task, and BP was recorded immediately at the end. A 5-min rest period (Baseline 2) came next, with HR, SC, and EMG recorded for the full period and BP recorded at the end. All subsequent baseline periods followed this format.

The neutral audiovisual presentation began with audiotaped instructions that described the procedure and asked the participant to practice making subjective ratings with the joystick. Next, each of the six neutral scenes was presented for 1 min, during which HR, SC, and EMG were recorded. Subjective units of distress (SUD) ratings were recorded at the end of each scene by having participants position a computer-generated arrow along a line that reflected the amount of distress they felt, from no distress to as much distress as they could imagine. The position of the arrow was translated to a numeric value ranging from 0 (*no distress*) to 100 (*the most that could be imagined*). A 5-min rest period (Baseline 3) followed the ratings for the final neutral scene. The combat audiovisual presentation and associated data collection were identical to the neutral presentation except for the content of the visual display and soundtrack. It was followed by a 5-min rest period (Baseline 4).

Tape-recorded instructions for the imagery procedure were delivered next, after which participants again rested quietly for 5 min (Baseline 5). Each script presentation consisted of four sequential 30-s periods: base rest, reading, imagining, and recovery. Briefly, the participant was told that four scripts would be presented and that he should listen carefully as each was read and vividly imagine it as though it were actually occurring. When the reading ended, he was to continue imagining the scene as vividly as possible, from beginning to end, until a tone was sounded. He was then to stop imagining the script and sit quietly until a second tone sounded (at the end of the recovery period). The participant then made ratings using a 12-point Likert-type scale about four dimensions of the experience: image vividness, perceived arousal, subjective pleasantness, and sense of control. The order of the four ratings varied across the scenes. The baseline for the next script began after 1 min or after HR had returned to within 5% of its value during the previous baseline period, whichever was longer.

Neutral scripts were presented first and third in the series, alternating with the scripts describing each participant's combat experiences. HR, SC, and EMG were recorded throughout the 2 min of each script sequence. BP was recorded only once, following the presentation of the fourth script. A 5-min rest period (Baseline 6) concluded the procedure.

The study clinician then provided all participants who were not scheduled for retesting with a thorough debriefing regarding the study and their experience in it. Psychophysiological retesting was scheduled for 303 participants (24% of the total) an average of 9.6 days later. These participants had their debriefing postponed until after the retest session or, for the 76 participants who terminated before retesting occurred, had debriefing provided by telephone.

Response definitions. We identified and replaced HR data artifact using a custom-designed statistical process.[1] SC values were identified as invalid and eliminated (i.e., set to missing) if they fell below an absolute value of 0.30 μS or if they reached a constant level at the upper or lower limit of the adjustable 16-μS range for recording. EMG values were identified as invalid if they fell below an absolute value of 0.25 μV or were part of a record that had a resting level exceeding 50 μV. The loss of BP data was infrequent and generally was the result of recording equipment failure.

[1] The absence of conventions for detecting and eliminating HR artifact led to a custom-designed approach to this task. The aim was to remove nonbiological variation while preserving as much of the biologically meaningful variation as possible. The method involved two routines from the S-Plus (Statistical Sciences, 1993) statistical package to recognize and adjust (i.e., "smooth") outlying values among the 6,240 in each complete participant record. Specifically, the HR data array was treated as a regularly spaced time series involving values collected at 0.5-s intervals for the 52 min of recording per session. The ar.gm routine was applied to these values to generate initial robust autoregression parameter estimates, which were then used by the robust smoother function, acm.ave. The smoothing process was applied twice, first to the raw data and then to the data that already had been smoothed once. A third application of the smoothing process was not found to markedly improve the outcome. This smoothing method was proposed by Martin (1981) and is based on a generalized M estimate for autoregression using a Tukey-type psi function. It has the attractive characteristic that data points determined to be nonartifactual (i.e., close to the signal) are left unaltered and gross outliers are replaced by estimates on the basis of the remainder of the time series. The net result of smoothing is minimization of bias and reduction of variability due to outlying values. Copies of the smoothing algorithm are available on request.

Baseline physiological scores for HR, SC, and EMG were based on samples averaged over all nonmissing values recorded during the final 2 min of the respective 5-min baseline periods. SBP and DBP were derived from the single reading recorded at the end of each baseline period. Mean physiological scores for task periods were calculated for successive 30-s intervals. This resulted in 4 intervals for the arithmetic stressor, 12 intervals each for the neutral and combat audiovisual presentations (i.e., 2 intervals per slide), and 4 intervals for each script presentation. HR, SC, and EMG samples were averaged across all nonmissing values within a 30-s interval. SBP and DBP were based on the single reading recorded at the end of each task period.

Response scores for HR, SC, and EMG during the arithmetic stressor were calculated by identifying the 30-s interval with the highest mean value and then subtracting the score for Baseline 1. SBP and DBP response scores were calculated by subtracting the Baseline 1 value from the arithmetic stressor value. HR, SC, and EMG response scores for the audiovisual presentations were calculated by subtracting the 30-s interval with the highest mean value during the neutral scenes from the 30-s interval with the highest mean value during the combat scenes, respectively, for each measure. SBP and DBP response scores were calculated by subtracting combat presentation values from those for the neutral presentation.

Response scores for HR, SC, and EMG during imagery scripts began with calculation of a mean value for the 30-s imagining periods within the two combat scripts and a comparable mean value for imagining periods within the two neutral scripts. The neutral script mean was subtracted from the combat script mean to create the index of response for each measure. SBP and DBP response scores for the script procedure were calculated by subtracting the respective values recorded at the end of Baseline 5 from those recorded following the final script presentation.

SUD ratings for the two audiovisual tasks were selected in a manner comparable with that used for the continuous physiological measures. The highest value from each slide presentation sequence was identified, and a response score was calculated by subtracting the highest value for a neutral scene from the highest value for a combat scene. Ratings for image vividness, perceived arousal, subjective pleasantness, and sense of control for the imagery scripts were calculated in a similar manner. For each, the average rating for the two neutral scripts was subtracted from the average rating for the two combat scripts to create the response index.

Data analysis. The Cooperative Study Program Coordinating Center provided centralized data management and analysis. Initial descriptive analysis examined demographic information, symptom reports, trauma exposure variables, comorbid diagnoses, and physiological variables from baseline and generic stressor periods. The second analysis examined physiological differences predicted between current and never PTSD groups, as well as tested for possible differences between the lifetime PTSD group and the other two groups. Two formats of analysis were used: (a) Continuous variables were subjected to an omnibus test of equality of means through an analysis of variance followed by post hoc pairwise comparisons based on the Tukey test and (b) categorical variables were compared across the three groups using a chi-square, with pairwise group comparisons from a Pearson's chi-square with Bonferroni adjustment of probability values. These analyses included the total sample of 1,328 participants who completed diagnostic interviewing.

Predictive modeling involved the application of logistic regression to develop an equation based exclusively on psychophysiological variables to classify participants in terms of their PTSD diagnostic status. Logistic regression analysis was used for this classification because it does not require that each predictor variable have a normal distribution of values.

Data from the 1,241 participants who began the first psychophysiological test session were used to the maximal degree allowed by each analytic procedure, given the pattern of individual missing values. The potential for classification accuracy associated with the logistic analysis was maximized by restricting participants to the current PTSD and never PTSD groups. Direct evaluation of model stability was accomplished by randomly dividing the sample into a calibration subsample, on which the model was derived, and a validation subsample, to which the model was then applied.

A stepwise procedure with adjustment for classification bias was used to fit the logistic regression model to the calibration sample. The resulting logistic model included coefficients b_i weighing each variable x_i, and these weights were used to generate a predictive score for each participant that reflected the probability that he belonged in the current PTSD group. Collectively, these probability values were compared with a set of classification values reflecting sensitivity, specificity, overall correct classification rate, and so forth. The logistic model developed from the calibration sample was used to determine the initial sensitivity and specificity of the classification procedure. The overall accuracy of the model was determined by examining the percentage of the validation sample that it correctly classified.

Follow-up analysis attempted to identify factors and influences that might account for prediction failures. Participants in the current PTSD group were rank ordered in terms of the logistic probability score calculated to predict PTSD group membership, and the respective 120 participants with the highest (responders) and lowest scores (nonresponders) were grouped. Comparisons were then conducted on a set of variables selected to address social, occupational, legal, military, trauma-related, general psychopathology, and personality domains.

Results

Group Comparisons

The set of demographic variables reflects enduring characteristics (e.g., race), as well as indexes of life functioning. As can be seen in Table 1, only current marital status, number of months in Vietnam, and years of postmilitary education are not different between current PTSD and both of the other groups. Differences indicate that the current PTSD group reported younger age and less education on arrival in Vietnam, more jobs and arrests since discharge, and less income. More individuals in the current PTSD group received disability income and were ethnically Hispanic. Finally, the current PTSD group also reported younger current age, more marriages, and Marine Corps service relative to the never PTSD group.

Psychopathology and combat exposure. The core set of measures for PTSD, comorbid psychopathology, and combat exposure are presented in Table 2. All measures show significant differences between the current and never PTSD groups, and 6 out of 10 show significant differences between the current and lifetime PTSD groups. The current PTSD group scored highest on the Mississippi Scale and the MMPI–2 PK scale indicators of PTSD; reported highest levels of exposure to combat; and had the highest rates of current comorbid diagnoses for major depression, panic disorder, and borderline personality disorder. In addition, the current PTSD group was higher than the never PTSD group with respect to rates of alcohol abuse or dependence and antisocial personality disorder. Over 85% of this latter group reported combat events that qualify as traumatic.

Psychophysiological measures. Initial baseline values and response scores for the arithmetic stressor are presented in Table 3. HR showed a clear pattern of baseline differences, with current PTSD higher than the other two groups. SC showed a similar pattern, although the current versus never PTSD difference did not reach significance. The other measures failed to even hint at group differences for Baseline 1. All three cardio-

Table 1
Comparison of Demographic Variables for Groups Based on PTSD Status

Demographic variable	Current *M*	Current *SD*	Lifetime *M*	Lifetime *SD*	Never *M*	Never *SD*	ANOVA *F* or χ^2	*p*<	Group difference
Current age (years)	43.2	3.1	43.6	4.0	45.0	4.9	F = 28.3	.001	C, L < N
Annual income ($, in thousands)	13.1	13.9	19.3	18.2	24.6	21.1	F = 59.5	.001	C < L < N
No. of jobs since military	25.0	26.6	16.0	18.9	12.1	16.4	F = 42.0	.001	C > L, N
No. of arrests since military	7.7	13.7	5.0	9.7	2.8	5.9	F = 23.1	.001	C > L > N
Age on arrival in Vietnam	19.8	2.5	20.6	3.4	21.7	4.3	F = 42.2	.001	C < L < N
No. of months in Vietnam	13.9	6.9	13.4	5.7	13.5	8.0	F < 1.0	*ns*	
Years of premilitary education	11.5	1.5	11.8	1.5	12.5	2.0	F = 42.4	.001	C < L < N
Years of postmilitary education	2.0	1.9	2.0	1.9	2.1	2.0	F < 1.0	*ns*	
Hispanic (% yes)	11		5		7		χ^2 = 9.4	.01	C > L, N
Currently married (% yes)	53		54		48		χ^2 = 2.8	*ns*	
Married more than once (% yes)	42		41		28		χ^2 = 22.2	.002	C, L > N
Receiving disability income (% yes)	48		29		27		χ^2 = 55.5	.001	C > L, N
Branch of military (% Marines)	27		23		15		χ^2 = 21.8	.001	C > N

Note. Sample size ranged from 771 to 773 for the current PTSD group (C), equaled 181 for the lifetime PTSD group (L), and ranged from 368 to 369 for the never PTSD group (N). Degrees of freedom for the analysis of variance (ANOVA) were 2 and 1317 or higher, *dfs* for the chi-square were 2, with a minimum sample size of 1,320. Group differences were determined by Tukey tests for continuous variables or by a chi-square for frequencies. PTSD = posttraumatic stress disorder.

vascular measures (HR, SBP, and DBP) showed a pattern of change during mental arithmetic that indicates less response in the current PTSD group compared with the other two groups, although only HR and DBP produced statistically reliable differences.

The predicted greater physiological response to trauma-related audiovisual presentations and imagery scripts for individuals with PTSD relative to those without PTSD is a key element of the study. Consistent with our prediction, Table 4 shows 7 out of 10 comparisons with greater response for the current PTSD group compared with the never PTSD group. HR, SC, and DBP showed significant differences across both task formats, whereas EMG differences were limited to the imagery procedure.

A unique dimension of this study was the inclusion of a lifetime PTSD group. As seen in Table 4, significant differences between the current PTSD group and the lifetime PTSD group were limited to HR and SC during the audiovisual presentation. By contrast, HR and EMG during the imagery scripts showed greater response for the lifetime PTSD group compared with the never PTSD group. Most of the remaining measures that did not show differences involving the lifetime PTSD group place the lifetime PTSD group intermediate between the other two groups.

Logistic Regression Analyses

Derivation of the logistic equation within the calibration subsample was based on stepwise regression (SAS Institute, 1990).

Table 2
Comparison of Psychometric Scales, Combat Exposure, and Comorbid Diagnoses for Groups Based on PTSD Status

Variable	Current *M*	Current *SD*	Lifetime *M*	Lifetime *SD*	Never *M*	Never *SD*	ANOVA *F* or χ^2	*p*<	Group difference
Mississippi Scale (total score)	123.2	18.5	96.8	18.4	78.7	20.6	F = 675.3	.001	C > L > N
MMPI–2 PK scale (raw score total)	30.3	9.5	18.8	9.2	13.5	10.4	F = 391.6	.001	C > L > N
Combat Exposure Scale (total score)	29.0	8.4	25.3	9.6	19.0	11.0	F = 136.4	.001	C > L > N
Vietnam combat-related Criterion A for PTSD diagnosis (% threshold)	100		99		86		χ^2 = 130.2	.001	C, L > N
Major depression (%)	36.2		10.0		6.2		χ^2 = 146.0	.001	C > L, N
Panic disorder (%)	12.9		3.9		0.3		χ^2 = 58.1	.001	C > L > N
Alcohol abuse or dependence (%)	23.8		18.8		16.3		χ^2 = 9.2	.01	C > N
Substance abuse or dependence (%)	12.1		13.3		8.1		χ^2 = 4.9	.09	
Antisocial personality disorder (%)	10.8		8.8		5.1		χ^2 = 9.8	.01	C > N
Borderline personality disorder (%)	17.9		4.4		3.5		χ^2 = 60.1	.001	C > L, N

Note. Sample size ranged from 729 to 778 for the current PTSD group (C), from 167 to 181 for the lifetime PTSD group (L), and from 349 to 369 for the never PTSD group (N). Degrees of freedom for the analysis of variance (ANOVA) were 2 and 1245 or higher, *dfs* for the chi-square were 2, with a sample size of 1,328 for all variables except PTSD Criterion A, which had 1,251 participants. Group comparisons were based on Tukey tests for continuous variables and a chi-square for frequencies. PTSD = posttraumatic stress disorder; MMPI–2 = Minnesota Multiphasic Personality Inventory; PK = Keane PTSD.

Table 3
Comparison of Physiological Baseline Levels and Response to a Generic (Arithmetic) Stressor for Groups Based on PTSD Status

Variable	Current *M*	Current *SD*	Lifetime *M*	Lifetime *SD*	Never *M*	Never *SD*	ANOVA *F*	*p*<	Group difference
Baseline level									
Heart rate (beats/min)	74.0	11.0	70.7	11.0	70.5	11.7	13.4	.001	C > L, N
Skin conductance (μS)	3.8	3.6	2.8	2.5	3.2	2.9	8.2	.001	C > L
Electromyogram (μV)	3.1	3.0	3.2	4.9	2.9	2.5	<1.0	*ns*	
Systolic blood pressure (mm)	120.2	14.6	120.0	14.7	120.0	14.3	<1.0	*ns*	
Diastolic blood pressure (mm)	71.7	8.9	71.0	8.3	71.1	8.9	<1.0	*ns*	
Response to arithmetic stressor									
Heart rate (beats/min)	6.2	5.2	8.1	6.2	7.5	5.2	11.3	.001	C < L, N
Skin conductance (μS)	2.4	2.1	2.5	2.3	2.6	2.3	<1.0	*ns*	
Electromyogram (μV)	3.7	5.3	3.7	4.7	3.3	4.4	<1.0	*ns*	
Systolic blood pressure (mm)	6.0	10.6	7.1	12.5	7.9	11.4	3.2	.05	
Diastolic blood pressure (mm)	4.9	5.5	6.1	6.2	6.5	5.6	9.0	.001	C < L, N

Note. Sample size ranged from 634 to 672 for the current PTSD group (C), from 144 to 154 for the lifetime PTSD group (L), and from 319 to 340 for the never PTSD group (N). Degrees of freedom for the analysis of variance (ANOVA) were 2 and 1098 or higher. Group differences were determined by Tukey tests. PTSD = posttraumatic stress disorder.

It identified the set from among 13 physiological variables that significantly and independently predict current PTSD group membership. As a second step, the regression analysis was reapplied to the calibration sample, with predictors restricted to the 4 identified as significantly contributing to the prediction equation. This second analysis was performed because logistic regression requires casewise deletion when a missing value is encountered on any variable in the array of potential predictors. Thus, some participants were not included in the initial stepwise analysis despite having a complete set of data for the 4 variables in the final model. For this second analysis, participants numbered 398 in the current PTSD group and 205 in the never PTSD group.

The logistic model that best predicted current PTSD group membership included HR level for Baseline 1 (X1), HR response to the audiovisual cues (X2), EMG response to the scripts (X3), and SC response to the scripts (X4). In mathematical notation, this model can be written as Log(Y/1 − Y) =

Table 4
Comparison of Physiological and Subjective Responses to Standardized (Audiovisual) and Idiographic (Imagery) Combat Stressors for Groups Based on PTSD Status

Variable	Current *M*	Current *SD*	Lifetime *M*	Lifetime *SD*	Never *M*	Never *SD*	ANOVA *F*	*p*>	Group difference
Audiovisual presentation									
Heart rate (beats/min)	2.4	5.3	0.9	4.8	0.5	4.6	17.6	.001	C > L, N
Skin conductance (μS)	0.8	1.5	0.5	1.2	0.3	1.0	17.2	.001	C > L, N
Electromyogram (μV)	1.0	3.9	0.3	2.3	0.3	3.1	5.7	.01	
Systolic blood pressure (μm)	3.7	10.2	2.0	8.3	2.5	7.5	3.1	.05	
Diastolic blood pressure (μm)	2.7	6.1	1.9	4.3	1.3	4.9	7.6	.001	C > N
Subjective units of distress	57.9	31.0	48.6	31.0	37.9	29.2	53.6	.001	C > L > N
Imagery script									
Heart rate (beats/min)	3.2	4.0	2.6	3.6	1.9	3.1	13.9	.001	C, L > N
Skin conductance (μS)	0.8	1.0	0.6	0.8	0.5	0.8	10.0	.001	C > N
Electromyogram (μV)	1.2	2.4	1.2	2.8	0.4	1.3	13.1	.001	C, L > N
Systolic blood pressure (μm)	2.9	9.8	2.3	9.1	1.0	8.8	4.4	.015	
Diastolic blood pressure (μm)	3.3	6.0	2.6	5.4	1.9	4.8	7.1	.001	C > N
(Un)pleasantness rating	6.0	3.5	5.2	3.6	5.4	3.1	5.6	.01	C > L

Note. Sample size ranged from 631 to 672 for the current PTSD group (C), from 144 to 154 for the lifetime PTSD group (L), and from 319 to 340 for the never PTSD group (N). Degrees of freedom for the analysis of variance (ANOVA) were 2 and 1096 or higher. Group differences were determined by Tukey tests. PTSD = posttraumatic stress disorder.

$-2.1555 + 0.0326\ (X1) + 0.0641\ (X2) + 0.2050\ (X3) + 0.4468\ (X4)$, where Y represents the probability that a participant will be a member of the current PTSD group. The Hosmer and Lemeshow (1989) test yielded a chi-square value of 9.2 ($df = 8$, $p = .32$), which indicates that the model fits well in terms of predicting PTSD diagnostic group membership. The crossover point for sensitivity and specificity is a probability value of .62, and the highest overall percentage for correct classification is 69%. Additionally, as can be seen in the top half of Table 5, there is a sharp trade-off between sensitivity and specificity at different probability levels.

The validation subsample included 200 participants from the current PTSD group and 103 participants from the never PTSD group, each of whom had values for all four variables in the logistic model. The equation applied to this sample produced values listed in the bottom half of Table 5. The crossover probability value for sensitivity and specificity was .67, and the highest overall percentage for correct classification was 64%. The range of values and trade-off between sensitivity and specificity was comparable with that obtained for the calibration subsample.

The logistic model from the calibration sample was then applied to the 139 participants in the lifetime PTSD group to further examine the group relationship to the current and never PTSD groups on psychophysiological response. The mean probability score for the entire lifetime group was .64 (range = .32 to .99). If this value is used as a cutoff for classification of the lifetime group, it results in an even split such that half are classified as current PTSD and half as never PTSD.

Finally, the logistic model was used to test the stability of key physiological measures across repeated tests. This was accomplished by generating PTSD probability scores for the 178 retest participants, drawn from all three PTSD groups, who had data available for all four variables in the model across both test sessions. Comparison of initial and retest probability scores ($M = .66$, $SD = .15$ and $M = .62$, $SD = .14$, respectively) reveals a significant reduction $t(177) = 3.57$, $p < .001$, and a Pearson correlation coefficient of .60 ($p < .001$) indicates a moderate relationship between probability scores across sessions.

Follow-Up Analysis of Logistic Equation Performance

The sensitivity and specificity findings indicate that an equation based exclusively on psychophysiological variables is moderately successful at classifying participants into groups established on the basis of the SCID. This level of performance is somewhat lower than that found in previous studies and invites further analysis to identify variables to account for individual differences in classification accuracy. Therefore, comparisons were conducted within the current PTSD group between the responder and nonresponder subgroups.

The majority of variables were not different between the responder and nonresponder groups. This included age (in Vietnam, at first trauma, and at PTSD onset), plus educational, marital (current status and number of marriages), occupational (number of jobs and income), and legal (arrests and convictions) histories. Likewise, exposure to potentially traumatizing events (before, during, and after the military), 14 out of 17 current PTSD symptoms from the SCID, and 10 of 11 current comorbid diagnostic conditions were not different between the groups. Finally, the preponderance of MMPI–2 validity indicators and clinical scales were equivalent for the groups, as was the proportion that tested positive for substances (both <9%).

As seen in Table 6, two clusters of variables from the analyses show differences between the groups. The first cluster reflects severity of PTSD such that responders scored higher on the Mississippi Scale and received higher ratings of PTSD severity from the study clinician and reported less nightly sleep, less sense of control during the idiographic trauma scripts, and more combat exposure. Higher proportions of responders reported intense distress on exposure to trauma reminders (SCID Item B4) and received service-connected disability. In addition, responders showed nonsignificant but consistent trends with re-

Table 5
Logistic Regression Classification Results (as Percentages) for PTSD Diagnosis Based on Physiological Variables

PTSD probability	Correct classification	Sensitivity	Specificity	False positive	False negative
		Calibration sample			
.55	69	83	42	27	44
.60	68	74	56	24	48
.65	64	63	66	22	52
.70	58	49	76	20	57
.75	52	36	84	19	60
		Validation sample			
.55	64	81	31	30	54
.60	63	74	44	28	54
.65	61	64	55	26	56
.70	58	52	71	23	57
.75	54	41	82	19	59

Note. The calibration sample included 398 from the current PTSD group and 205 from the never PTSD group. The validation sample included 200 from the current PTSD group and 103 from the never PTSD group. PTSD = posttraumatic stress disorder.

Table 6
Differences in PTSD Severity, Guilt, and Depression Between Physiological Responders and Nonresponders Within the Current PTSD Group

Variable	Responder *M*	Responder *SD*	Nonresponder *M*	Nonresponder *SD*	*t* test or χ^2	*p*<
Mississippi Scale (total score)	127.1	18.0	119.4	19.2	$t = 3.1$	.01
Combat Exposure Scale (total score)	30.8	8.6	27.6	9.8	$t = 2.6$	.02
Clinician rating of PTSD severity (SCID)	2.2	0.7	2.0	0.8	$t = 2.2$	.05
Global Assessment of Functioning (SCID)	50.7	9.0	53.1	10.3	$t = 1.8$	.08
Reported hours of sleep per night	4.8	1.6	5.3	1.6	$t = 2.4$	.02
Combat-neutral subjective rating for audiovisual presentation	60.7	31.4	53.3	32.6	$t = 1.8$	.02
Trauma-neutral rating of control for imagery scripts	5.7	3.9	4.4	4.2	$t = 2.4$	.08
Laufer-Parsons Inventory (total score)	101.8	31.1	87.3	28.3	$t = 3.6$	.001
MMPI-2 Depression Scale (T score)	81.8	13.6	76.5	15.0	$t = 2.8$	.01
Receives service-connected disability (% yes)	60		47		$\chi^2 = 4.8$	.05
SCID Item B4: "Intense psychological distress at exposure to events" (% threshold)	78		66		$\chi^2 = 4.0$	.05
SCID Item D6: "Physiologic reactivity upon exposure to events" (% threshold)	80		70		$\chi^2 = 3.2$	.08
SCID Item C7: "Sense of foreshortened future" (% threshold)	71		59		$\chi^2 = 3.6$	.001

Note. Sample size ranged from 106 to 120 for the responder group and from 100 to 120 for the nonresponder group. Degrees of freedom ranged from 211 to 238 for *t* tests and were adjusted for unequal variances as necessary; the chi-square had 1 degree of freedom and a sample size of 240 for each variable. PTSD = posttraumatic stress disorder; SCID = Structured Clinical Interview for *DSM-III-R;* MMPI-2 = Minnesota Multiphasic Personality Inventory.

spect to SUD ratings for the audiovisual combat presentation, as well as with respect to ratings by the study clinician in terms of subjective physiological response on exposure to trauma reminders (SCID Item D6) and general functioning. Overall, the responders showed more severe PTSD symptoms and poorer functioning.

The second pattern relates to guilt and depression. In this instance, the LPI and the Depression Scale from the MMPI-2 were the primary differentiating variables, with the SCID item concerning "sense of foreshortened future" showing a similar trend. Examination of the LPI items (e.g., "[I] should have died in the war" and "[I get] upset because [I] feel a buddy or comrade got killed because of something [I] did or did not do") reveals themes of dysphoria and self-recrimination that characterize the responses of the responder group.

Finally, there is a substantial difference between groups with respect to racial composition, with the responder group being 80% Caucasian compared with 55% for the nonresponder group, $\chi^2(1, N = 240) = 17.1, p < .001$). This finding raises the prospect that differences between groups may be due to race. This possibility was addressed by analysis of covariance, controlling for Caucasian status, applied to all continuous variables that differentiated the groups, and by similarly adjusted chi-square analysis applied to the categorical variables. Some variables showed a modest decrease in the magnitude of group difference after adjustment for race, but key distinguishing variables (e.g., the Mississippi Scale, the Combat Exposure Scale, and the LPI) did not.

Discussion

Results of this multisite clinical trial provide definitive support for a positive association between psychophysiological responsivity to cues depicting traumatic war-zone experiences and combat-related PTSD. The physiological differences included higher resting levels shown by the current PTSD group during the initial baseline and greater response to both standardized audiovisual presentations and idiographic imagery scripts that contained trauma-relevant cues.

Differences between PTSD and non-PTSD groups at baseline for HR and SC replicated previous findings (Blanchard, Kolb, Gerardi, Ryan, & Pallmeyer, 1986; Pitman et al., 1987). Responses to mental arithmetic indicate that the participants in all of the groups were broadly equal in their reaction to cues that were generically threatening or challenging (see also Blanchard, Kolb, Taylor, & Wittrock, 1989). Importantly, analyses within the current PTSD group show that nonresponders were as responsive as responders when the cues were not trauma relevant. Accordingly, we favor the view that baseline differences reflect acute anticipatory fear to challenge testing by individuals with PTSD (see Prins et al., 1995), although the present design does not exclude the possibility of other acute or tonic influences on arousal.

The between-groups response differences were also consistent with previous studies but were of smaller magnitude than expected. For example, HR response in the current PTSD group averaged less than 3 beats/min compared with reported values of 5–10 beats/min (e.g., Blanchard et al., 1982; Malloy et al., 1983; Pitman et al., 1987). Follow-up within the current PTSD group demonstrated that the most responsive individuals showed average HR changes in the expected 6–7 beats/min range. Because they are ascribed more severe symptoms of PTSD by clinician interviewers and endorse considerably more symptoms of war-related guilt and depression, this pattern indicates that individuals with severe PTSD are more responsive to challenge testing. In turn, this leads to consideration of the possibility that previous psychophysiological studies enrolled participants with more severe PTSD.

Comparison of Tables 1 and 2 with comparable demographic and psychometric information from previous trials indicates that the PTSD participants in the present study had lower Mississippi Scale and PK scale scores, more stable marital histories, and fewer comorbid diagnoses than participants in prior studies (see Keane et al., 1988; Keane & Wolfe, 1990; Malloy et al., 1983). In addition, there were clear differences in characteristics of comparison groups. Almost without exception, prior investigations enrolled non-PTSD controls with superior social adjustment and less psychopathology than the present never PTSD group.

Classification rates based on psychophysiological response were lower than those reported previously. Although some of this reduction can be attributed to a sampling strategy that was broader than prior investigations, it is also likely that classification findings were influenced by the lower psychophysiological responding shown by the current PTSD group. Still, the approximately two-thirds rate of correspondence between psychophysiological measures and current PTSD diagnostic status is a respectable and promising performance for a biological test of a psychological condition, surpassing classification observed in relation to other biological tests (e.g., the dexamethasone suppression test; Insel & Goodwin, 1983).

It seems clear that some individuals exposed to Criterion A events can report a constellation of symptoms that qualify for a PTSD diagnosis and can demonstrate help seeking, social impairment, and vocational dysfunction and yet not manifest physiological reactivity to trauma-related cues. In addition, there are a variety of self-protective or self-regulatory maneuvers that might be used by individuals to disengage from the trauma-related challenge task or otherwise dampen physiological response (see Orr & Kaloupek, 1997). Such maneuvers may account for some instances of nonresponse by individuals with PTSD. However, a broader perspective on the issue recognizes the possibility that psychophysiological responding to trauma-relevant cues is relevant to assessment or treatment by virtue of its imperfect correspondence with interview-based PTSD diagnosis. Differences in clinical picture and treatment prognosis may emerge in relation to this marker (see Blanchard et al., 1996), as well as relationships to possible subtypes of the disorder (Quinn, Kaloupek, Keane, Hsieh, & Lavori, 1996).

This study contained a large number of participants who met criteria for a lifetime diagnosis of PTSD. These individuals showed physiological responsivity ranging between that shown by current and never PTSD groups. The protocol also incorporated a test of the stability of the psychophysiological measures over two sessions separated by an average of nearly 10 days. Findings indicate a statistically significant reduction in PTSD probability scores on retesting, coupled with moderately strong Pearson correlation coefficients for the scores across sessions. This pattern indicates that there is some systematic loss of responsiveness, although the level of consistency remains reasonably high in comparison with other test–retest examinations of psychophysiological responding (e.g., Arena, Blanchard, Andrasik, Cotch, & Myers, 1983; Tomarken, 1995).

In conclusion, the present study provides strong empirical support for the presence of objectively measured psychophysiological reactivity to trauma cues as a nomothetic-distinguishing feature of PTSD. The findings suggest that individuals with the strongest physiological responses are the most impaired on clinician- and self-rating scales and endorse more symptoms of war-related guilt and depression. Physiologically based classification rates in the study were 69% and 64% in the calibration and validation samples, respectively. Finally, findings from an array of psychological tests, questionnaires, and interviews indicate that Vietnam veterans with current PTSD experience a broad range of psychological symptoms, multiple comorbid conditions, marital and familial dysfunction, vocational impairment, financial instability, and punitive involvement with the law, all of which justify further efforts to understand, measure, and treat this substantial public health problem.

References

Arena, J. G., Blanchard, E. B., Andrasik, F., Cotch, P. A., & Myers, P. E. (1983). Reliability of psychophysiological assessment. *Behavior Research and Therapy, 21,* 1–13.

Blanchard, E. B., Hickling, E. J., Buckley, T. C., Taylor, A. E., Vollmer, A., & Loos, W. R. (1996). Psychophysiology of posttraumatic stress disorder related to motor vehicle accidents: Replication and extension. *Journal of Consulting and Clinical Psychology, 64,* 742–751.

Blanchard, E. B., Kolb, L. C., Gerardi, R. J., Ryan, P., & Pallmeyer, T. P. (1986). Cardiac response to relevant stimuli as an adjunctive tool for diagnosing post-traumatic stress disorder in Vietnam veterans. *Behavior Therapy, 17,* 592–606.

Blanchard, E. B., Kolb, L. C., Pallmeyer, T. P., & Gerardi, R. J. (1982). A psychophysiological study of post-traumatic stress disorder in Vietnam veterans. *Psychiatric Quarterly, 54,* 220–229.

Blanchard, E. B., Kolb, L. C., & Prins, A. (1991). Psychophysiological responses in the diagnosis of posttraumatic stress disorder in Vietnam veterans. *Journal of Nervous and Mental Disease, 179,* 99–103.

Blanchard, E. B., Kolb, L. C., Taylor, A., & Wittrock, D. (1989). Cardiac response to relevant stimuli as an adjunct in diagnosing post-traumatic stress disorder: Replication and extension. *Behavior Therapy, 20,* 535–543.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1990). *Minnesota Multiphasic Personality Inventory (MMPI–2): Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

Dobbs, D., & Wilson, W. P. (1960). Observations on persistence of war neurosis. *Diseases of the Nervous System, 21,* 686–691.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18,* 232–239.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23,* 567–589.

Gerardi, R., Keane, T. M., & Penk, W. (1989). Utility: Sensitivity and specificity in developing diagnostic tests of combat-related post-traumatic stress disorder (PTSD). *Journal of Clinical Psychology, 45,* 691–703.

Hosmer, D. W., Jr., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Insel, T. R., & Goodwin, F. K. (1983). The dexamethasone suppression test: Promises and problems of diagnostic laboratory tests in psychiatry. *Hospital and Community Psychiatry, 34,* 1131–1138.

Kardiner, A. (1941). *The traumatic neuroses of war.* New York: Hoeber.

Keane, T. M., Caddell, J. M., & Taylor, K. L. (1988). Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology, 56,* 85–90.

Keane, T. M., Fairbank, J. A., Caddell, J. M., Zimering, R. T., Taylor, K. L., & Mora, C. A. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment, 1,* 53–55.

Keane, T. M., Malloy, P. F., & Fairbank, J. A. (1984). Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 52,* 888–891.

Keane, T. M., & Wolfe, J. (1990). Comorbidity in post-traumatic stress disorder: An analysis of community and clinical studies. *Journal of Applied Social Psychology, 20,* 1776–1788.

Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines.* Newbury Park, CA: Sage.

Lang, P. J., Levin, D. N., Miller, G. A., & Kozak, M. J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective response integration. *Journal of Abnormal Psychology, 92,* 276–306.

Laufer, R. S., Yager, T., Frey-Wouters, E., & Donnellan, J. (1981). *Legacies of Vietnam: Volume III. Post-war trauma: Social and psychological problems of Vietnam veterans and their peers.* Washington, DC: U.S. Government Printing Office.

Malloy, P. F., Fairbank, J. A., & Keane, T. M. (1983). Validation of a multimethod assessment of posttraumatic stress disorders in Vietnam veterans. *Journal of Consulting and Clinical Psychology, 51,* 488–494.

Martin, R. D. (1981). Robust methods for time series. In D. F. Findley (Ed.), *Applied time series analysis* (pp. 683–759). New York: Academic Press.

Orr, S. P., & Kaloupek, D. G. (1997). Psychophysiological assessment of PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 69–97). New York: Guilford Press.

Orr, S. P., Pitman, R. K., Lasko, N. B., & Herz, L. R. (1993). Psychophysiologic assessment of posttraumatic stress disorder imagery in World War II and Korean combat veterans. *Journal of Abnormal Psychology, 102,* 152–159.

Pitman, R. K., Orr, S. P., Forgue, D. F., de Jong, J. B., & Claiborn, J. M. (1987). Psychophysiologic assessment of posttraumatic stress disorder imagery in Vietnam combat veterans. *Archives of General Psychiatry, 44,* 970–975.

Prins, A., Kaloupek, D. G., & Keane, T. M. (1995). Psychophysiological evidence for autonomic arousal and startle in traumatized adult populations. In M. J. Friedman, D. S. Charney, & A. Y. Deutch (Eds.), *Neurobiological and clinical consequences of stress: From normal adaptation to PTSD* (pp. 291–314). New York: Raven Press.

Quinn, S. J., Kaloupek, D. G., Keane, T. M., Hsieh, F. Y., & Lavori, P. W. (1996, November). *Heterogeneity in posttraumatic stress disorder diagnosis from the DVA Cooperative Study #334.* Paper presented at the 12th annual meeting of the International Society for Traumatic Stress Studies, San Francisco, CA.

Rosenheck, R., & Fontana, A. (1989). *War Stress Interview (WSI).* West Haven, CT: Department of Veterans Affairs, Northeast Program Evaluation Center.

SAS Institute. (1990). *Proc Logistic* (Version 6.10). Carey, NC: Author.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. D. (1989). *Structured Clinical Interview for DSM-III-R—Patient Version (SCID-P).* New York: New York State Psychiatric Institute.

Statistical Sciences. (1993). *S-Plus for Windows reference manual* (Version 3.1). Seattle, WA: Author.

Tomarken, A. J. (1995). A psychometric perspective on psychophysiological measures. *Psychological Assessment, 7,* 387–395.

Wenger, M. A. (1948). Studies of autonomic balance in Army air forces personnel. *Comparative Psychology Monographs, 19*(4).

Received December 3, 1996
Revision received November 23, 1997
Accepted May 8, 1998 ■