Development and Initial Validation of a Statistical Prediction Instrument for Assessing Combat-Related Posttraumatic Stress Disorder

Brian P. Marx, PhD,† Kathryn L. Humphreys, Ed.M,* Frank W. Weathers, PhD,‡ Elaine K. Martin, BS,* Denise M. Sloan, PhD,*† William M. Grove, PhD,§ Danny G. Kaloupek, PhD,*† and Terence M. Keane, PhD*†*

Abstract: Statistical analyses were used to derive and validate a statistical prediction instrument to determine combat-related posttraumatic stress disorder (PTSD) status. Participants were 1081 Vietnam veterans with and without combat-related PTSD. The statistical prediction instrument, which consisted of 12 well-known risk and resilience variables associated with PTSD, proved to be an accurate and efficient means of detecting PTSD among participants and compared well against other existing self-report measures of PTSD. The instrument's practical applications and its use in clinical appraisals of PTSD are discussed.

Key Words: Posttraumatic stress disorder, statistical prediction, assessment, diagnosis, veterans.

(*J Nerv Ment Dis* 2008;196: 605–611)

Mental health professionals have developed a number of standardized assessment instruments to detect the presence of combat-related posttraumatic stress disorder (PTSD) (see Keane et al., 2004, for a review). Keane et al. (1987) recommended that an assessment for PTSD is best achieved through the use of multiple reliable and valid instruments, because every measure is associated with some degree of error. This multimethod approach combines data derived from self-report measures, structured clinical interviews and, when possible, psychophysiologic assessment. Such multimodal assessment of PTSD takes advantage of each measure's relative strengths, overcoming the psychometric limitations of any single instrument and maximizing correct diagnostic decisions.

Although a wide variety of methods have been used in the assessment of PTSD, one promising method that has not been explored to date is the use of statistical prediction instruments (SPIs) (Grove and Meehl, 1996; Meehl, 1954; Swets et al., 2000). SPIs use statistical analyses of cases with known outcomes to determine which risk factors are relevant to a given diagnostic decision and to what extent. An SPI accepts case-based values of the variables and combines them to give the best possible assessment or summary of the available evidence. Such instruments have potential value from a public healthcare perspective, in terms of both screening for the disorder and promoting more efficient allocation of resources and efforts to those with the greatest need for services.

Using a statistical prediction approach to determine PTSD caseness would capitalize on the fact that extensive information is now available regarding the risk and resilience factors that influence long-term adjustment to trauma exposure (e.g., King et al., 1999; Ozer et al., 2003; Schnurr et al., 2004). Empirically identified predictors include quality of early family environment, age at trauma exposure, history of prior adversity, severity of trauma exposure, exposure to atrocities-abusive violence, peritraumatic responses, breadth and strength of the social support network, exposure to additional life stressors, and individual difference variables such as coping repertoires, hardiness, and neurobiology. The development of an SPI would extend this PTSD risk-resilience research by creating an assessment device using information about the known PTSD risk-resilience factors.

The statistical prediction approach would also be a valuable complement to the most widely used existing methods of assessing PTSD (i.e., diagnostic interviews, self-report) because it would not require individuals to report directly about their subjective PTSD symptoms. Much of the information for an SPI would still be collected via self-report, but the link between items and a PTSD diagnosis would be more indirect and less obvious, and thus the SPI would have considerably less face validity. This would reduce the extent to which response bias, especially symptom over-reporting and malingering, interferes with obtaining a valid assessment of PTSD status.

Because SPIs are based on the premise that risk factors may be appropriately quantified to aid in diagnostic decision making, perhaps the ultimate utility of an effective SPI would

*Department of Psychiatry, National Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts; †Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts; ‡Department of Psychology, Auburn University, Auburn, Alabama; and §Department of Psychology, University of Minnesota, Minneapolis, Minnesota.

Send reprint requests to Brian P. Marx, PhD, Behavioral Science Division, National Center for PTSD, 150 S. Huntington Ave. (116B-2), Boston, MA 02130. E-mail: brian.marx@va.gov.

ISSN: 0022-3018/08/19608-0605
DOI: 10.1097/NMD.0b013e318181326d

be in detecting those who may be prone to develop PTSD following trauma before symptoms actually manifest. Such a tool would be useful from a prevention perspective and would reduce overall healthcare utilization by PTSD patients.

This article describes a proof of concept study establishing that an SPI can be developed and used to accurately and efficiently ascertain military-related PTSD status among male veterans. We hypothesized that a statistical prediction measure of PTSD would demonstrate strong diagnostic utility for screening and differential diagnosis, and would compare favorably in this regard with other well-established self-report measures of PTSD. We also identified optimal cut scores for the SPI and explored the association of the SPI with measures of functional impairment and response validity.

METHOD

Data Source

Data were taken from the responses of 1081 veterans who participated in VA Cooperative Study #334 (CS334; see Keane et al., 1998). Participants in the study were veterans who served in the Vietnam theater of operations between August 1964 and May 1975 and who were currently using VA services during their participation. Recruitment took place over a 42-month period from inpatient and outpatient programs in psychiatry, substance abuse, and PTSD.

For the purposes of this investigation, we included data only from those participants who were either diagnosed with current PTSD or never diagnosed with PTSD. Participants were administered a comprehensive assessment battery including structured interviews and self-report measures. They supplied information on a broad array of topics related to pre-war background and functioning, military and war-zone experiences, and post-war circumstances, life events and mental health status. Drawing on the findings of King et al. (1999), we chose to focus on those risk-resilience factors that were found to have the strongest relations with combat-related PTSD status. Specifically, we examined report of early trauma history, age at entry into Vietnam, exposure to atrocities-abusive violence, self-reported postmilitary trauma, and structural and functional indicators of social support. We did not include indicators of perceived threat and malevolent environment in this investigation because these factors were not adequately assessed in CS334. In line with other previous studies, we included additional salient risk-resilience factors in this investigation. These factors include trauma-related guilt (Henning and Frueh, 1997; Kubany et al., 1995), peritraumatic dissociation (Kaufman et al., 2002), and extent of combat exposure (Dohrenwend et al., 2006).

Measures

Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Third Edition-Revised (DSM-III-R) (SCID) (Spitzer et al., 1989)

The SCID is a structured diagnostic interview that assesses disorders in the DSM-III-R (American Psychiatric Association, 1987). SCID interviews were audiotaped and 128 were rated independently by a second clinician. Interrater reliability for PTSD diagnostic categories current, lifetime, or never was 77% ($\kappa = 0.68$, with weights of 0, 0.5, and 1.0). An additional subset of 36 interviews rated by the second clinician was reviewed again. The interrater reliability for these 36 interviews was 78% ($\kappa = 0.66$, with weights of 0, 0.5, and 1.0).

Combat Exposure Scale (Keane et al., 1989)

This seven-item questionnaire assesses the presence and severity of wartime combat experiences. All 7 items were individually examined in the initial analyses to determine which risk-resilience variables best discriminated between individuals with and without PTSD.

Laufer-Parsons Inventory (LPI) (Laufer et al., 1981)

This 33-item measure assesses combat-related guilt. Questions refer to the participant's feelings of guilt over the past 6 months. We included the LPI because previous studies have shown that guilt during or immediately following trauma is a strong predictor of PTSD (Kubany et al., 1995). Further, some have suggested that guilt related to perpetrating or observing atrocities may be an important risk factor for combat-related PTSD (Beckham et al., 1998; McNally, 2003; Yehuda et al., 1992). We chose the 5 items most highly correlated ($r = 0.80$ or higher) with the LPI total score to use in our initial analyses to determine which risk-resilience variables best discriminated between individuals with and without PTSD.

War Stress Interview (Rosenheck and Fontana, 1989)

This questionnaire was used to assess a number of demographic features, experiences during deployment, and general and psychological functioning.

Mississippi Scale (MS) for Combat-Related PTSD (Keane et al., 1988)

The MS is a 35-item, widely used and extensively validated self-report measure of combat-related PTSD. The MS was included to compare its diagnostic performance with that of the SPI.

Minnesota Multiphasic Personality Inventory-2 (MMPI-2) (Butcher et al., 1989)

The MMPI-2 is a 567-item multiscale inventory that assesses personality and psychopathology, and is one of the most widely used psychological assessment instruments. For this study, we included the PTSD Keane (PK) scale (Keane et al., 1984), which in the MMPI-2 consists of 46 items found to discriminate combat veterans with and without PTSD, to compare its diagnostic performance with that of the SPI. We also included several MMPI-2 validity scales, Lie (L), Infrequency (F), Correction (K), Infrequency Psychiatric Inpatient (Fp), Infrequency PTSD (Fptsd) (Arbisi and Ben-Porath, 1995; Elhai et al., 2002), in this investigation to assess the association between the PTSD measures used in this study and various forms of response bias.

SPI Development

We divided the entire sample into 3 equal subsamples matched on PTSD diagnosis, either current or never PTSD as known from the SCID I, as well as PTSD symptom severity calculated using MS scores. These 3 subsamples did not differ significantly in age, education, or ethnicity. Two of these subsamples were used to derive and identify variables that significantly differentiate between individuals with and without PTSD. Items that best discriminated individuals with and without PTSD were identified using one-way ANOVAs. In order for a variable to be included in the SPI, the variable must have significantly predicted PTSD diagnostic status, as well as attained a medium to large effect size in both derivation subsamples. Specifically, continuously scored items were retained for the SPI if the Ω^2 value for the difference between the PTSD and non-PTSD sample was 0.12 or greater in both derivation subsamples (see Olejnik and Algina, 2003). Similarly, dichotomously scored items were retained if their risk difference scores were 0.25 or greater in both derivation subsamples (risk difference is an effect size index used when assessing risk and recommended over other indices; see Rosenthal, 2001).

A total of 12 variables met or exceed the selection criterion (see Table 1). We then derived empirically based cutoffs for each item that was not already dichotomous by using the risk difference statistic to determine which cutoff point best distinguished between those with and without PTSD. In each case, the cutoff point with the highest risk difference in each derivation subsample was used in the final scale. The risk difference scores were all 0.30 or greater.

Once the cutoffs for each variable were identified, we considered the best method for computing a total score for the Posttraumatic Stress Disorder SPI (PSPI, from this point onward). Although some have suggested that compared with a unitary weighting approach an optimal weighting approach to items may enhance accuracy of prediction (e.g., Dawes et al., 1989; Swets et al., 2000; Wainer, 1976), we decided to employ a unitary weighting approach for several reasons. First, from a clinical utility standpoint, we thought that a measure that did not require a complicated total score calculation would be more appealing to clinicians. Second, other well known and widely used psychological measures, such as the MMPI-2 and Wechsler Adult Intelligence Scale, also use a unitary weighting approach. Finally, the expected loss in predictability that might result from using the unitary weighting approach would be less than 2.5% under all but the most highly implausible assumptions about the distribution of the optimal regression weights (W. M. Grove, personal communication, April 12, 2007). Thus, participants were initially awarded one point for each item they endorsed in the direction of PTSD.

Participants in the Derivation Sample

The derivation sample consisted of 721 participants, two-thirds of the original sample. The mean age of participants at the time of the study was 43.09 (*SD* = 3.60) and the mean number of years of education was 13.84 (*SD* = 2.43). Participants were primarily White (67%), African-American (19%), and Hispanic (9%), and over half (54%) were currently married. Sixty-seven percent were diagnosed with PTSD, and 43% had a service-connected disability.

Descriptives

Summing the PSPI items yielded a total score ranging from 0 to 12, with higher scores indicating greater likelihood of having combat-related PTSD. Because all 12 items met or exceeded the selection criterion in both derivation subsamples, the following analyses were conducted on the combined subsamples. The PSPI displayed strong internal consistency (Cronbach's alpha = 0.84), with interitem cor-

TABLE 1. PSPI Items and Frequencies of Endorsement (%) for the PTSD and Non-PTSD Groups

PSPI Item with Cutoff Points Specified		PTSD Group	Non-PTSD Group
1	Observed or participated in atrocities (% yes)	76.3	32.1
2	Sit with anyone dying from battle wounds (% yes)	64.5	34.2
3	Number of times fired rounds at your enemy (3 times or more)	90.1	56.0
4	Number of times saw someone hit by incoming or outgoing rounds (3 times or more)	80.6	43.8
5	Overwhelming feelings of guilt about the war (sometimes or more often)	77.4	21.3
6	Feelings that your conscience bears a heavy burden for actions during the war (sometimes or more often)	75.2	14.0
7	Thoughts that the things you did in war were unforgivable (almost never or more often)	69.1	16.1
8	Thoughts that nobody could love you for what you did in the war (almost never or more often)	70.1	16.2
9	Feelings that you do not deserve the good things in life (sometimes or more often) during most upsetting event during the war:	66.9	12.8
10	Reported to have "blanked out" or "spaced out"	56.4	23.3
11	Reported to have experienced a sense of time change	84.1	58.3
12	Reported to have been surprised after the event by finding out things had happen at the time that you were not aware of	69.5	37.9

PTSD, Posttraumatic Stress Disorder; Non-PTSD, Individuals without Posttraumatic Stress Disorder; PSPI, PTSD Statistical Prediction Instrument.

relations ranging from 0.08– 0.61. The mean score on the PSPI for individuals with PTSD was 8.85 ($SD = 2.40$) and the mean score for non-PTSD participants was 3.64 ($SD = 2.47$). A *t*-test revealed that the groups differed significantly on the PSPI total score, $t(431.83) = -26.11$, $p < 0.001$). Table 1 displays the prevalence of each group in their endorsement of the PSPI items. In all cases, the directionality of each item was arranged so that the PTSD group's endorsement was greater than that of the non-PTSD group.

Diagnostic Utility

Signal detection analyses (Kraemer, 1992) were conducted to compare the diagnostic utility of the MS, PK scale, and PSPI for predicting current PTSD diagnostic status based on the SCID PTSD module. According to Kraemer (1992), the optimally efficient cutoff is the score that yields the highest quality of efficiency or $\kappa(.5)$, and is best for differential diagnosis. The optimally sensitive cutoff is the score that yields the highest quality of sensitivity or $\kappa(1)$, and is best for screening. Optimal cutoffs for PTSD diagnosis for each measure were determined by dichotomizing it at every possible cutoff and examining the quality of efficiency and sensitivity.

Information regarding the optimally efficient cutoffs is shown in Table 2. The MS at a cutoff of 100 had the highest quality of efficiency (0.88), followed by the PSPI at a cutoff of 6 (0.87), and the PK scale T score at a cutoff of 64 (0.83). At these cutoffs all 3 measures had excellent sensitivity, and for the MS and PSPI this was combined with good specificity. The PK scale, however, had substantially lower specificity than did the other 2 measures, and this accounts for its lower quality of efficiency. Information regarding the optimally sensitive cutoffs is shown in Table 3. Each of the 3 measures displayed nearly perfect sensitivity at their optimally sensitive cutoff, and for the MS and PSPI this was combined with moderate specificity. Again, however, the PK scale had substantially lower specificity than did the other 2 measures. These findings indicate that the PK scale would be less effective as a screener because it would yield higher levels of false positives than would the MS or the PSPI.

The MS had the strongest point-biserial correlation with SCID PTSD diagnosis ($r = 0.74$), followed by the PSPI ($r = 0.71$), and the PK scale ($r = 0.64$). Total scores on the PSPI were significantly related to scores on the MS ($r = 0.73$) and PK scale ($r = 0.62$), as well as negatively correlated with clinician Global Assessment of Functioning ratings ($r = -0.50$). Because the PSPI does not require individuals to report directly about their PTSD symptoms it may serve to reduce the extent to which symptom under- and over-reporting interfere with obtaining a clear picture of PTSD status. Bolstering this possibility is the finding that PSPI scores were significantly less correlated with MMPI-2 validity (fake-good and fake-bad) scales than were the MS and PK scale (see Table 4).

RESULTS

Participants in the Cross-Validation Sample

The cross-validation sample consisted of 360 participants, the remaining third of the original sample. The mean age of participants at the time of the study was 43.46 ($SD = 4.17$) and the mean number of years of education was 13.96 ($SD = 2.44$). Again, the participants were primarily White (65%), African-American (20%), and Hispanic (11%), and less than half (46%) were married at the time of the study. Over two-thirds (68%) were diagnosed with PTSD, and 39% had a service-connected disability.

Using the optimal cutoffs identified in the derivation sample, we analyzed the diagnostic utility of the PSPI, MS, and PK scale in the cross-validation sample. As shown in Table 5, the performance of all 3 tests in the cross-validation

TABLE 2. Optimally Efficient Cutoff Scores for the Mississippi Scale, PK Scale T Score, and PSPI in the Full Derivation Sample

Test	*n*	*p*	Q	Sens	Spec	PPV	NPV	Eff	$\kappa(0)$	$\kappa(0.5)$	$\kappa(1)$
Mississippi scale >100	721	0.67	0.66	0.90	0.85	0.92	0.80	0.88	0.77	0.74	0.71
PK scale T score >64	700	0.67	0.71	0.90	0.68	0.85	0.76	0.83	0.55	0.59	0.65
PSPI >6	683	0.67	0.67	0.90	0.80	0.90	0.80	0.87	0.70	0.70	0.71

PTSD diagnostic status based on Structured Clinical Interview for DSM-III-R (SCID).

PTSD, Posttraumatic Stress Disorder; PK Scale, Keane PTSD Scale; PTSD, Posttraumatic Stress Disorder; PSPI, PTSD Statistical Prediction Instrument; *p*, prevalence of PTSD diagnosis; q, level of test (proportion of test positives); Sens, sensitivity, Spec, specificity, PPV, positive predictive value, NPV, negative predictive value; $\kappa(0)$, kappa coefficient representing quality of specificity; $\kappa(0.5)$, kappa coefficient representing quality of efficiency; $\kappa(1)$, kappa coefficient representing quality of sensitivity.

TABLE 3. Optimally Sensitive Cutoff Scores for the Mississippi Scale, PK Scale T Score, and PSPI in the Full Derivation Sample

Test	*n*	*p*	q	Sens	Spec	PPV	NPV	Eff	$\kappa(0)$	$\kappa(0.5)$	$\kappa(1)$
Mississippi scale >80	721	0.67	0.82	0.99	0.54	0.82	0.96	0.84	0.44	0.60	0.94
PK scale T score >48	700	0.67	0.88	0.99	0.34	0.75	0.94	0.78	0.25	0.40	0.91
PSPI >3	683	0.67	0.88	0.99	0.36	0.76	0.94	0.78	0.27	0.41	0.91

PTSD diagnostic status based on Structured Clinical Interview for DSM-III-R (SCID).

PTSD, Posttraumatic Stress Disorder; PK Scale, Keane PTSD Scale; PSPI, PTSD Statistical Prediction Instrument; *p,* prevalence of PTSD diagnosis; q, level of test (proportion of test positives); Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; $\kappa(0)$, kappa coefficient representing quality of specificity; $\kappa(0.5)$, kappa coefficient representing quality of efficiency; $\kappa(1)$, kappa coefficient representing quality of sensitivity.

sample was consistent with their performance in the validation sample, with respect to the previously derived optimally efficient cutoffs. Sensitivity and specificity dropped slightly for the PSPI, specificity dropped slightly for the PK scale, and the performance of the MS was unchanged.

As shown in Table 6, similar results were found for the previously derived optimally sensitive cutoffs. Again, the MS performed nearly identically in this sample as the derivation sample. The PK scale and PSPI continued to display excellent sensitivity, and each test's specificity improved in this sample, with the PSPI's specificity rising by 0.07.

Next, we examined the point-biserial correlations between the 3 measures and SCID diagnosis in the cross-validation sample. Again, the MS had the strongest point-biserial correlation with SCID PTSD diagnosis ($r = 0.74$), followed closely by the PSPI ($r = 0.73$), and then the PK scale ($r = 0.64$). The correlation between the PSPI and SCID diagnosis improved in the cross-validation sample compared with the original derivation sample, while the other measures correlation remained the same.

Table 7 presents the PSPI, MS, and PK scale correlations with MMPI-2 validity scales from the MMPI-2. The PSPI was significantly less correlated with these scales than both other measures, with the exception of the Fptsd scale, as it did not significantly differ from the MS scale. Across both samples, the PSPI was less strongly correlated to these validity indices than the MS and PK scale.

TABLE 4. Correlations with MMPI-2 Validity Scales in the Full Derivation Sample

Variable	PSPI	Mississippi Scale (MS)	PK Scale (PK)	Group Difference
L	−0.20*	−0.24*	−0.33*	PSPI, MS < PK
F	−0.52*	0.71*	0.85*	PSPI < MS < PK
K	−0.45*	−0.61*	−0.75*	PSPI < MS < PK
Fb	0.52*	0.73*	0.86*	PSPI < MS < PK
F_p	0.31*	0.41*	0.50*	PSPI < MS < PK
F_{ptsd}	0.22*	0.29*	0.39*	PSPI < MS < PK

Group differences were calculated using Williams's T2 statistic, recommended by Steiger (1980) for use in testing the difference between dependent correlations, *$p < 0.05$ was used to define a significant group difference.

PTSD, Posttraumatic Stress Disorder; MMPI-2, Minnesota Multiphasic Personality Inventory-2; PSPI, PTSD Statistical Prediction Instrument; PK Scale, Keane PTSD Scale; L, MMPI-2 Lie Scale; F, MMPI-2 Infrequency Scale; K, MMPI-2 Defensiveness Scale; Fb, MMPI-2 Back Side F scale; $F_{p,}$ MMPI-2 Infrequency-Psychopathology Scale; $F_{ptsd,}$ MMPI-2 Infrequency-Posttraumatic Stress Disorder Scale.

DISCUSSION

Consistent with our predictions, the results from this study suggest that it is feasible to construct a reliable and valid SPI for current military-related PTSD using known information about the relative effects of pre-military, warzone, and post-military risk factors on the development of PTSD. Because the DSM diagnostic criteria for PTSD require exposure to a potentially life threatening experience during which the individual experiences specific responses, the development and validation of a statistically-informed diagnostic aide that assesses traumatic exposure, peritraumatic responses, and other risk-resilience factors and then uses that information to determine a likelihood of PTSD gives researchers and clinicians another empirically-based method for detecting the presence of PTSD, or even a means of identifying those who may be vulnerable to developing the disorder following traumatic exposure.

The SPI created in this investigation displayed strong diagnostic utility for screening and differential diagnosis. Further, with respect to diagnostic utility and brevity (12 items versus 35 for the MS and 46 for the PK scale), the PSPI compared favorably to other already well-established self-report measures of PTSD. Results from this study suggest that quantifying risk and resilience factors to determine the likelihood of having PTSD may be another viable method to

TABLE 5. Optimally Efficient Cutoff Scores From the Derivation Sample Applied in the Validation Sample

Test	*n*	*p*	Q	Sens	Spec	PPV	NPV	Eff	κ(0)	κ(0.5)	κ(1)
Mississippi scale >100	360	0.68	0.66	0.90	0.85	0.92	0.80	0.88	0.77	0.74	0.71
PK scale T score >64	349	0.67	0.71	0.90	0.66	0.84	0.77	0.82	0.53	0.59	0.66
PSPI >6	345	0.67	0.65	0.86	0.77	0.88	0.73	0.83	0.64	0.62	0.60

PTSD diagnostic status based on Structured Clinical Interview for DSM-III-R (SCID).

PTSD, Posttraumatic Stress DisorderPK Scale = Keane PTSD Scale; PSPI = PTSD Statistical Prediction Instrument; p = prevalence of PTSD diagnosis; q = level of test (proportion of test positives); Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value; κ(0) = kappa coefficient representing quality of specificity; κ(0.5) = kappa coefficient representing quality of efficiency; κ(1) = kappa coefficient representing quality of sensitivity.

TABLE 6. Optimally Sensitive Cutoff Scores From the Derivation Sample Applied in the Validation Sample

Test	*n*	*p*	q	Sens	Spec	PPV	NPV	Eff	κ(0)	κ(0.5)	κ(1)
Mississippi scale >80	360	0.68	0.82	0.99	0.54	0.82	0.95	0.84	0.43	0.59	0.93
PK Scale T score >48	349	0.67	0.88	0.98	0.33	0.75	0.90	0.77	0.24	0.37	0.86
PSPI >3	345	0.67	0.84	0.98	0.43	0.78	0.93	0.80	0.33	0.48	0.89

PTSD diagnostic status based on Structured Clinical Interview for DSM-III-R (SCID).

PTSD, Posttraumatic Stress Disorder; PK Scale, Keane PTSD Scale; PSPI, PTSD Statistical Prediction Instrument; p, prevalence of PTSD diagnosis; q, level of test (proportion of test positives); Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; κ(0), kappa coefficient representing quality of specificity; κ(0.5), kappa coefficient representing quality of efficiency; κ(1), kappa coefficient representing quality of sensitivity.

TABLE 7. Correlations with MMPI-2 Validity Scales in the Validation Sample

Variable	PSPI	Mississippi Scale (MS)	PK Scale (PK)	Group Difference
L	−0.19*	−0.27*	−0.38*	PSPI < MS < PK
F	0.57*	0.71*	0.84*	PSPI < MS < PK
K	−0.50*	−0.66*	−0.80*	PSPI < MS < PK
Fb	0.59*	0.75*	0.87*	PSPI < MS < PK
F_p	0.31*	0.40*	0.46*	PSPI < MS, PK
F_{ptsd}	0.22*	0.28*	0.31*	PSPI < PK

Group differences were calculated using Williams's T2 statistic, recommended by Steiger (1980) for use in testing the difference between dependent correlations, $*p < 0.05$ was used to define a significant group difference.

PTSD, Posttraumatic Stress Disorder; MMPI-2, Minnesota Multiphasic Personality Inventory-2; PSPI, PTSD Statistical Prediction Instrument; PK Scale, Keane PTSD Scale; L, MMPI-2 Lie Scale; F, MMPI-2 Infrequency Scale; K, MMPI-2 Defensiveness Scale; Fb, MMPI-2 Back Side F scale; F_p, MMPI-2 Infrequency-Psychopathology Scale; F_{ptsd}, MMPI-2 Infrequency-Posttraumatic Stress Disorder Scale.

incorporate into a comprehensive assessment battery for evaluating PTSD status among military personnel and veterans.

Importantly, results of this investigation showed that the optimally efficient cutoff for the PSPI showed some shrinkage in the validation study. This finding suggests that we may not yet have the most optimal set of items for discriminating those with and without PTSD using an SPI. On the other hand, the optimally sensitive cutoff on the PSPI held up well in the cross-validation sample and suggests that the measure developed here may serve reasonably well as a PTSD screener. Future research may be needed to address these questions. Of course, an important limitation of the measure developed here is that it was developed using cross-sectional, retrospective data. More accurate predictions and discriminations between diagnostic groups require longitudinal data.

Since Meehl's seminal work on statistical prediction in 1954, research has shown that a properly developed and applied mechanical data combining method is likely to help in diagnosing and predicting human behavior, yielding decisions and predictions on average that are distinctly more accurate than the clinical method, even when the judge has access to more information (Grove et al., 2000). Even when such methods are only as accurate as the clinical method, the algorithm is faster and cheaper.

When mechanical methods are demonstrated to be more accurate than clinical judgment (the usual scenario), the benefits to patients (and to associated stakeholders and to society in general) are more apparent and stronger. Further, for predictive tasks in which there is significant superiority of the formula, utility, and ethical considerations dictate that the decision to use actuarial methods should not be difficult.

Despite the overwhelming evidence in favor of the actuarial method as well as the fact that such methods are less expensive, most professionals continue to use a subjective, clinical judgment approach when making predictions or diagnostic decisions. According to Grove and Meehl (1996), probably the 2 most important reasons why clinicians do not use mechanical prediction methods are (1) that they are not well educated about what they are and how they can be used to optimize clinical decision making, and (2) ready-to-use decision aids are not known to them for their often-encountered prediction problems. Hopefully, demonstrations such as this one showing that SPIs can be developed and easily implemented may both educate others and provide a necessary example of a viable decision aide.

Generally speaking, mechanical data combination methods can be made as immune to major pitfalls of unreliable or nonoptimal data aggregation as the sample size used to build them will allow. They also do not suffer from the cognitive and decision making biases repletely documented in the cognitive literature. Further, use of a mechanical combining method has some tendency to vitiate problems with discrepancies between raters (criterion variability manifesting as interrater un-reliability). This is because the mechanical summing of variables (in typical statistical formulas) or its equivalent in other algorithmic but nonlinear methods of data combination tends to cause rater criterion variance, as a form of systematic error in one predictor variable, to average with other errors-in-variables and so to cancel out when variables are combined. As such, they may be used to improve diagnostic performance by improving diagnostic accuracy.

In the case of PTSD, given the limitations of the existing methods which are used to determine diagnostic status (e.g., subjective self-report) as well as clinicians' difficulties in optimally to aggregating assessment data, there would potentially be a good deal to gain from using algorithms to detect or even screen for the disorder. Specifically, developing and using algorithms based on traumatic exposure and risk-resilience variables may promote more accurate diagnoses of PTSD as well as enhance the utility of the diagnoses for stakeholders (most particularly for the principal stakeholder, the patient, but also for others including the clinicians who serve patients, and the taxpayers who fund the care patients receive) by promoting fair compensation and access to care for those with the disorder in primary care, forensic, emergency (psychiatric triage) and even military theater of operations settings.

Although PTSD is a significant problem among traumatized individuals, the majority of those exposed to war-zone trauma are either resilient (in that the event causes no significant impairment or functional disruption) or recover on their own after a period of disruption (e.g., Dohrenwend et al., 2006). Despite this, we currently have no available methods for accurately discriminating between those individuals who will develop military-related PTSD and those who will not. The ability to make distinctions between those more and less vulnerable to PTSD onset is critical for appropriate allocation of resources to early intervention and prevention efforts. SPIs for PTSD may allow clinicians and health professionals to more competently make such distinctions and target those individuals who may be at greatest risk for the disorder.

One of the more interesting findings from this study was that scores from the PSPI are less significantly correlated with measures of invalid responding than other self-report indices of PTSD. This suggests that the PSPI may be less affected by symptom exaggeration and dissimulation than other measures of PTSD. Given the longstanding concern that military-related PTSD is over diagnosed among service members and veterans (Burkett and Whitley, 1998; McHugh and Treisman, 2007) and that secondary gain may be a source of misdiagnosis, measures that are less vulnerable to misrepresentation or response bias may be a helpful addition to a comprehensive evaluation for PTSD.

The results of this study may be limited in their generalizability to all veterans of military service, as this sample is comprised of treatment-seeking Vietnam era veterans. Future research on the PSPI, or other assessment devices like it, should use a more heterogeneous sample to determine its broader relevance. Future studies should examine additional risk and resilience factors not assessed here from other official sources (e.g., military personnel records) and use nonlinear methods of data combination (e.g., Classification and Regression Trees, Neural Networks) and cross validation (e.g., bootstrapping) to enhance diagnostic accuracy and efficiency. Other studies should also use longitudinal research designs and examine whether or not an SPI may be developed to discriminate between those who may suffer from chronic PTSD from those who may ultimately recover on their own from the disorder.

CONCLUSIONS

The results of this study suggest that SPIs may compliment other methods used to determine PTSD status among traumatized individuals. The PSPI detected the presence of PTSD more accurately than the PK scale and nearly as well as the MS scale. Our findings suggest that this type of assessment method may be less vulnerable to symptom feigning or exaggeration. Although SPIs can be used as a detection instrument, using SPIs to identify those particularly at-risk for developing psychopathology is a promising next step.

REFERENCES

American Psychiatric Association (1987) *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed, revised).Washington (DC): Author.

Arbisi PA, Ben-Porath YS (1995) An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency–Psychopathology Scale, F(p). *Psychol Assessment*. 7:424–431.

Beckham JC, Feldman ME, Kirby AC (1998) Atrocities exposure in Vietnam combat veterans with chronic posttraumatic stress disorder: Relationship to combat exposure, symptom severity, guilt and interpersonal violence. *J Trauma Stress*. 11:777–785.

Burkett BG, Whitley G (1998) Stolen Valor: How the Vietnam Generation was Robbed of its Heroes and its History. Dallas (TX): Verity Press Publishing.

Butcher JN, Dalhstrom WG, Graham JR, Tellegen A, Kaemmer B (1989) *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual and Administration for Scoring.* Minneapolis (MN): University of Minnesota Press.

Dawes R, Faust D, Meehl P (1989) Clinical versus actuarial judgment. *Science*. 243:1668–1674.

Dohrenwend BP, Turner JB, Turse NA, Adams BG, Koenen KC, Marshall R (2006) The psychological risks of Vietnam for U.S. veterans: A revisit with new data and methods. *Science*. 313:979–982.

Elhai JD, Ruggiero KJ, Frueh BC, Beckham JC, Gold PB, Feldman ME (2002) The Infrequency-Posttraumatic Stress Disorder Scale (Fptsd) for the MMPI-2: Development and initial validation with veterans presenting with combat-related PTSD. *J Pers Assess*. 79:531–549.

Grove W, Meehl P (1996) Comparative efficacy of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychol Public Pol L*. 2:293–323.

Grove W, Zald D, Lebow B, Snitz B, Nelson C (2000) Clinical versus mechanical prediction: A meta-analysis. *Psychol Assessment*. 12:19–30.

Henning K, Frueh B (1997) Combat guilt and its relationship to PTSD symptoms. *J Clin Psychol*. 53:801–808.

Kaufman M, Kimble M, Kaloupek D, McTeague L, Bachrach P, Forti A, Keane T (2002) Peritraumatic dissociation and physiological response to trauma-relevant stimuli in Vietnam combat veterans with posttraumatic stress disorder. *J Nerv Ment Dis*. 190:167–174.

Keane TM, Caddell JM, Taylor KL (1988) Mississippi Scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *J Consult Clin Psych*. 56:85–90.

Keane TM, Fairbank JA, Caddell M, Zimering RT, Taylor KL, Mora CA (1989) Clinical evaluation of a measure to assess combat exposure. *Psychol Assessment*. 1:53–55.

Keane TM, Kolb LC, Kaloupek DG, Orr SP, Blanchard EB, Thomas RG, Hsieh FY, Lavori PW (1998) Utility of psychophysiological measurement in the diagnosis of posttraumatic stress disorder: Results from a Department of Veterans Affairs Cooperative Study. *J Consult Clin Psych*. 66:914–923.

Keane T, Malloy P, Fairbank J (1984) Empirical development of an MMPI subscale for the assessment of combat-related posttraumatic stress disorder. *J Consult Clin Psych*. 52:888–891.

Keane T, Street AE, Stafford A (2004) The assessment of military-related PTSD. In JP Wilson, TM Keane (Eds), *Assessing Psychological Trauma and PTSD* (pp 262–288). New York: Guilford.

Keane T, Wolfe J, Taylor K (1987) Post-traumatic stress disorder: Evidence for diagnostic validity and methods of psychological assessment. *J Clin Psychol*. 43:32–43.

King DW, King LA, Foy DW, Keane TM, Fairbank JA (1999) Posttraumatic stress disorder in a national sample of female and male Vietnam veterans: Risk factors, war-zone stressors and resilience-recovery variables. *J Abnorm Psychol*. 108:164–170.

Kraemer HC (1992) Evaluating medical tests: Objective and Quantitative Guidelines. London: Sage.

Kubany ES, Abueg FR, Owens JA, Brennan JM, Kaplan AS, Watson SB (1995) Initial examination of a multidimensional model of trauma-related guilt: Applications to combat veterans and battered women. *J Psychopathol Behav*. 17:353–376.

Laufer RS, Yager T, Frey-Wouters E, Donnellan J (1981) Legacies of Vietnam, Vol. III: Post-War Trauma: Social and Psychological Problems of Vietnam Veterans and Their Peers. Washington (DC): U.S. Government Printing Office.

McHugh P, Treisman G (2007) PTSD: A problematic diagnostic category. *J Anxiety Disord*. 21:211–222.

McNally R (2003) Progress and controversy in the study of posttraumatic stress disorder. *Annu Rev Psychol*. 54:229–252.

Meehl P (1954) Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. Minneapolis (MN): University of Minnesota Press.

Olejnik S, Algina J (2003) Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychol Methods*. 8:434–447.

Ozer E, Best S, Lipsey T, Weiss D (2003) Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychol Bull*. 129:52–73.

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

Rosenthal R (2001) Effect sizes in behavioral and biomedical research: Estimation and interpretation. In L Bickman (Ed), *Validity and Social Experimentation* (pp 121–139). Thousand Oaks (CA): Sage.

Schnurr PP, Lunney CA, Sengupta A (2004) Risk factors for the development versus maintenance of posttraumatic stress disorder. *J Trauma Stress*. 17:85–95.

Spitzer RL, Williams JBW, Gibbon M, First MD (1989) *Structured Clinical Interview for DSM-1II-R—Patient Version* (SCID-P). New York: New York State Psychiatric Institute.

Swets JA, Dawes RM, Monahan J (2000) Psychological science can improve diagnostic decisions. *Psychol Sci Pub Int*. 1:1–21.

Wainer H (1976) Robust statistics: A survey and some prescriptions. *J Educ Stat*. 1:285–312.

Yehuda R, Southwick SM, Giller EL (1992) Exposure to atrocities and severity of chronic posttraumatic stress disorder in Vietnam combat veterans. *Am J Psychiat*. 149:333–336.

