International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 15(2): 57–63 (2006)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.185

A comparative study of posttraumatic stress disorder assessment under standard conditions and in the field

CHRISTOPHER R. ERBES,[1,2] THOMAS N. DIKEL,[3] RAINA E. EBERLY,[4,5] WILLIAM F. PAGE,[6] BRIAN E. ENGDAHL[1,5]

1 Minneapolis Veterans Affairs Medical Center, Minneapolis, Minnesota, USA
2 Department of Psychiatry, University of Minnesota, Minnesota, USA
3 Private Practice, Gainesville, Florida, USA
4 Vet Center, St Paul, Minnesota, USA
5 Department of Psychology, University of Minnesota, USA
6 Medical Follow-up Agency, Institute of Medicine, National Academy of Sciences, USA

Abstract

Little is known about the performance of clinician-administered structured diagnostic interviews when given under variable levels of examiner training and monitoring. We sought to explore this question. We examined the performance of a self-report questionnaire and a structured clinical interview in the assessment of post-traumatic stress disorder (PTSD) in two community samples of war veterans. One sample was interviewed under standard conditions (N = 372) and the other under unknown and less standardized conditions (N = 420), more closely approximating 'field conditions'. Interview results were used to predict questionnaire-based PTSD status. Kappas, sensitivities, specificities, and positive predictive powers were moderate and of similar magnitude in both samples. Our results suggest that even under uncertain ('field') conditions, clinician-administered structured interviews can produce results comparable to those produced under more tightly controlled conditions. Copyright © *2006 John Wiley & Sons, Ltd.*

Key words: clinical assessment, structured interview, PTSD

Introduction

Diagnosis drives research into, and treatment of, psychiatric conditions, including PTSD. Diagnoses are used to decide which participants are included as 'cases' of PTSD in studies, and to determine which patients will (or will not) receive specific treatments for PTSD. Some argue that there is no single 'gold standard' for diagnosing PTSD (Schlenger et al., 2004) but there has been a large emphasis on the use of structured clinical interviews as diagnostic tools in both research and practice (Weathers et al., 2001; Weiss, 2004). Clinician-administered structured interviews are favoured because they combine opportunities for clinical judgement and interpretation with rigorous guidelines for assessing specific symptoms of a disorder. As such, structured clinical interviews attempt to optimize diagnostic agreement across clinicians and contexts.

Structured clinical interviews are typically developed under controlled conditions, then reported in the literature, and finally administered clinically in the field. Thus, reliability and validity data are derived in controlled settings where the examiners using the diagnostic interviews are carefully trained and closely monitored to ensure an adequate use of the interview. For example, in a study comparing the PTSD module of the Structured Clinical Interview for DSM-III-R

(SCID), the Clinician Administered PTSD Scale (CAPS), and the PTSD Symptom Scale – Interview (PSS-I), Foa and Tolin (2000) used 22 clinicians administering between 2 and 3 structured clinical interviews to each of 64 participants. The developers of the instruments trained each clinician and each interview was videotaped and reviewed to ensure appropriate use.

Clinical and research administration conditions often vary from the conditions reported in instrument development and validity studies. Structured interviews may be used by clinicians who receive little specific training or supervision with the structured interview. This presents a dilemma for the clinician and researcher. How closely do results obtained in these clinical or research settings resemble those intended by the developers of the interview? To what degree do structured clinical interview reliability and validity research data generalize to more general settings? Research is needed to assess the degree to which variation in interviewer training and interview context affect instrument validity (see, for example, Aziz and Kenford, 2004). As far as we are aware, previous research has not addressed this concern.

This study addresses this question by examining the performance of a structured clinical interview and a self-report instrument in two community samples of war veterans. Our review of the literature found no data directly addressing the performance of structured clinical interview in controlled versus 'field' conditions. Therefore, ours was an exploratory study. We examined the comparative validity of a clinician-administered structured interview for PTSD across controlled versus field conditions, as indexed by its relationship with a standard self-report PTSD questionnaire.

Subjects and methods

One sample was studied under controlled conditions, using the PTSD module of the SCID (Spitzer and Williams 1986), and the Mississippi Scale for Combat-Related PTSD (M-PTSD) (Keane et al., 1988). This 'standard condition' sample was evaluated under circumstances similar to those used in development and validation studies for structured clinical interviews. The SCID and the M-PTSD were also applied to a second sample under less standardized conditions, more closely approximating assessment 'field conditions.' Because the M-PTSD is a well validated self-report instrument that should not be affected by administration conditions, we expected it to perform similarly under the two conditions.

Instruments

The M-PTSD scale is a 35-item, Likert-scaled, self-report instrument originally developed to assess PTSD symptoms and various associated clinical features. In a sample of 362 American Vietnam War veterans, it yielded a high internal consistency (alpha = 0.94), sensitivity (0.93), and specificity (0.89). In a subsample of 39 veterans assessed over an approximately 7 day interval, the test-retest Pearson correlation was 0.97 (Keane et al., 1988). M-PTSD cutoff scores of 89–91 have demonstrated excellent convergence against clinician assessed PTSD in community samples. In a sample of 343 Vietnam veterans it exhibited agreement with a SCID PTSD diagnosis in 86% of cases (Kulka et al., 1991). In samples of American Korean War and WWII veterans, the M-PTSD showed agreement with the SCID in 85% of 325 cases (Engdahl et al., 1996) and 81% of 257 cases (Page, 1992).

The SCID has frequently been used to assess PTSD in research and clinical settings (for example, Kulka et al., 1991). It provides specific operational criteria for the 17 DSM-IV symptoms of PTSD within the re-experiencing, numbing/avoidance, and increased arousal criterion categories. The PTSD module of the SCID has been used as a 'gold standard' for evaluating the validity of other structured clinical interviews (Weiss, 2004). The SCID had a reported kappa of 0.93 reflecting high inter-rater agreement in a national sample of Vietnam veterans (Kulka et al., 1991).

Subjects and procedure

Two samples of community-residing World War II (WWII) and Korean War combat veterans and former prisoners of war (POWs) were studied in Minnesota (Engdahl et al. 1996) and nationally (Page, 1992).

Minnesota sample (standard condition)

Former prisoners of war were randomly selected from POW rosters based on military repatriation records. Mailings and follow-up telephone contacts with 344 potential participants resulted in 262 completed assessments (76%). Forty-four (13%) were willing to participate but could not be scheduled because of ill health or distance from the Medical Center; 38 (11%) declined to participate. Fifty-six (21%) were former POWs held by Japan, 191 (73%) by Germany, and 15 (6%) by North

DOI: 10.1002/mpr

Korea. These proportions matched national proportions, as did their average reported percentages of weight lost while held as a POW. Participants and non-participants had similar lengths of captivity, comparable marital and family status, and similar magnitudes of service-connected disability ratings.

The non-POW combat veterans responded to community announcements published in veteran-oriented periodicals. None were excluded, 110 provided complete data sets, and six participants who started the exam process did not complete it. The representativeness of this group is unknown as population data are not available. As noted below, however, PTSD is probably over-represented in this group.

All 372 participants were reimbursed for their participation. All were male and resided in Minnesota, Wisconsin, or North Dakota. Two were Native American, one was Hispanic, and the rest were white. Median age was 70 and median education was 12 years. Approximately two-thirds of the subjects were receiving at least some of their health care from a Veterans' Affairs Medical Center (VAMC); 7% were involved in mental health care at the time of recruitment.

Diagnostic interviews and psychological testing were completed at the VAMC in Minneapolis between August 1991 and August 1994. All participants were administered the full SCID Non-Patient module, the SCID PTSD module, and the M-PTSD scale by one of two doctoral level or one of two masters level psychologists experienced in PTSD assessment. Additional psychodiagnostic testing was completed with results reported elsewhere (Engdahl et al., 1996; Engdahl et al., 1997; Engdahl et al., 1998). All four interviewers viewed SCID training videotapes and followed a detailed guide for the SCID (Schlenger and Allison, 1987). They observed and rated each other's interviews. Eight interviews were directly observed by a second rater resulting in eight pairs of ratings. Five interviews were taped and independently reviewed by two additional raters, yielding three pairs of observations for each of these five participants, resulting in 15 pairs of ratings for these cases. There were no disagreements as to the presence or absence of current or lifetime PTSD among these 23 possible pairs of ratings.

National sample (field condition)

This consisted of randomly selected POWs and combat controls (matched by theatre of war) who had participated in a series of studies over the previous 40 years (summarized by Page, 1992). Examination invitations were mailed to 2044 POWs. After excluding 294 deceased subjects and 257 subjects who recently completed a similar examination, 52% (779/1493) completed at least some portion of the examination. The SCID PTSD modules were completed by 38% (295) of these POW participants. Examination invitations also were mailed to 1604 control subjects. After excluding 314 deceased subjects, 39% (508/1290) completed at least some portion of the examination and SCIDs were obtained for 28% (140) of the participating control subjects. Participants were not paid for completing examinations but were reimbursed for mileage expenses. Contrary to written requirements in the project's directive, examiners often failed to administer the SCID. We have no evidence suggesting that this occurred on a systematic basis. Many participants were given the M-PTSD to complete at home and did not return it, leading to missing data for this measure.

With one exception, participants were similar to non-participants in both the POW and control groups. Veterans' Association hospitalization rates, age, education, and marital status were similar; participants were somewhat more likely to have served in the Army Air Corps. Participants were 93.2% White, 4.4% Black, and 2.3% American Indian, Hispanic, or Asian. Median age was 68 and median education was 12 years.

The assessments were conducted between 1989 and 1991 at 88 US VAMCs. Nearly all were conducted at the VAMC closest to the veteran's home. Assessments typically required a full day to complete and included a medical examination, a social history interview, the SCID PTSD module, and the M-PTSD scale. All examiners also were asked to conduct a non-structured clinical examination to determine the presence of PTSD. The order of the interviews and test administration was not specified.

Each VAMC had a designated POW physician coordinator charged with overseeing the process. Some coordinators administered the SCIDs themselves but in the majority of cases they assigned this task to regular staff members or non-VA clinicians (compensation and pension consultants). In accord with VA policy, all examiners were required to be licensed for independent practice in their respective fields, indicating that they were at least moderately experienced diagnosticians. Most of these examiners presumably had previous experience in the clinical assessment of combat-related PTSD. Of the 435 SCIDs (295 POWs plus 140 controls),

DOI: 10.1002/mpr

60% were completed by examiners with MDs, 10% by PhDs, 4% by MSWs, and 1% by DOs; 25% of the SCIDs did not indicate the examiner's degree.

Veterans' Administration-wide SCID training had not been provided and it was the impression of the fourth author (WFP), who visited 32 of the 88 sites during the study, that few of these examiners had been trained in SCID administration. No such training was provided to these examiners as part of the present study. Several paragraphs concerning test administration were included in the directive sent to all the VAMCs. This directive may or may not have been available to the examiners. It described structured interviews in positive terms, noting they can lead to improved data collection and analysis. It encouraged examiners to use their own approach to establish rapport and provided language to introduce the SCID. It urged examiners to familiarize themselves with the flow and the sequence of the wordings. No specific instructions were given on scoring the module, but the rules are imbedded in the module itself.

Raw data from 435 SCIDs in the national sample were transcribed and entered by a subcontractor in 1991. Transcription and coding errors resulted in complete data sets for only 321 subjects (Page, 1992). As many as 114 SCIDs were therefore unusable due to missing data in this initial transcription process. As this is a study on the performance of the SCID, rather than of data transcription processes in research, we reviewed the raw data on the SCIDs (including the medical records documenting the assessments) in order to locate information that had been lost during the transcription process. By doing so the number of usable cases with complete SCID data was raised to 394. In 28 cases we located apparent examiner errors in SCID application (for example, incorrect application of scoring algorithms). Because such errors reflected a misapplication of the instrument by examiners, they were not corrected in the data set.

Results

The percentage of overall agreement between the M-PTSD and SCID diagnoses was 83.8% in the standard condition and 84.4% in the field condition. Kappa statistics were calculated within each sample to assess the match between PTSD diagnoses from the SCID and the M-PTSD. Based on previous research (Kulka et al., 1991; Page, 1992; Spiro et al., 1994), and prior findings in the Minnesota sample (Engdahl et al., 1996), a M-PTSD cutoff of 89 was used in both samples for predicting the presence of PTSD. Kappa for the standard sample was 0.58 (95% confidence interval = 0.49–0.68) and for the field sample was .57 (95% confidence interval = 0.47–0.66). Using the standard errors for each kappa, a test for significant differences between the two kappas was not significant (Z = 0.26). Using the PTSD classification from the M-PTSD as the standard for PTSD 'caseness', sensitivity, specificity, and positive and negative predictive power were calculated for each sample and are presented in Table 1. Differences in proportion of true positives (sensitivity), true negatives (specificity), and predictive power (positive and nega-

Table 1. Convergent validity statistics for SCID predictions of M-PTSD-based PTSD status under standard and field conditions

Sample	n	kappa	Sensitivity[a]		Specificity[b]		Positive predictive power[c]		Negative predictive power[d]	
				X^2		X^2		X^2		X^2
Standard	370	0.58	0.69	0.03	0.89	0.06	0.69	0.83	0.89	0.06
Field	385	0.57	0.71		0.88		0.63		0.88	

Note. SCID = Structured Clinical Interview for DSM-III-R, Post-traumatic Stress Disorder module. M-PTSD = Mississippi Scale for Combat-Related Post-traumatic Stress Disorder. PTSD = post-traumatic stress disorder. Standard = Minnesota sample. Field = national sample.

[a] = proportion of true PTSDs (based on the M-PTSD) correctly identified by the SCID.

[b] = proportion of true non-PTSDs (based on the M-PTSD) correctly identified by the SCID.

[c] = proportion of true cases (based on M-PTSD) among SCID-identified PTSDs.

[d] = proportion of true negatives (based on M-PTSD) among SCID-identified non-PTSDs.

DOI: 10.1002/mpr

tive) between samples were evaluated by constructing the four relevant 2 × 2 contingency tables and calculating a chi-square for each table. For example, the difference in sensitivity between the two samples was compared in a contingency table in which the number of true positives and false positives (which are the values from which specificity is derived) were put in a contingency table for both samples. Thus, each chi-square analysis represented a test for differences between the two samples in the specific proportions in question. Kappas, sensitivities, specificities, and positive predictive powers were moderate to moderately high and of similar magnitude in both samples. Chi-square analyses revealed no significant differences between the two conditions on these indices.

Although ours is not a study of PTSD rates per se, we present the current PTSD rates for both samples as assessed via the M-PTSD and the SCID in Table 2 for the POW groups, the combat groups, and the combined groups.

The similarity of rates across the POW groups (both drawn from community rosters) indicate that the SCID module functioned comparably in the two samples and further supports the comparability of SCID performance across conditions. The only significant difference was observed between the combat samples using the Mississippi (a trend towards a similar difference was present with the SCID). Combat controls were selected randomly for the national sample; however, in the Minnesota sample, combat controls had responded to a community announcement about a study of adjustment among combat veterans. The higher rates of PTSD in the Minnesota sample most likely reflect a self-selection factor: combat veterans with PTSD symptoms may have been more likely to volunteer for the study.

Discussion

This study compared the performance of a common structured clinical interview for PTSD (the PTSD module of the SCID) in two samples that had distinct evaluation conditions. The standard (Minnesota) condition employed assessments conducted in a single medical centre by doctoral and masters-level psychologists who were specifically trained in the use of the SCID and whose performance was monitored and evaluated through the use of videotaped interviews that were rated by multiple coders. The broader national sample (field condition), involved assessments taking place across 88 different VAs by a variety of examiners including physicians, psychologists, social workers, and osteopaths. Training in this broader sample was non-standardized and involved no specific or systematic training in the use of the SCID and no checks for reliability in interviewing or scoring. Despite the differences between these assessment conditions, the SCID performed similarly in terms of agreement with a standardized self-report measure of PTSD. The percentage of agreement between the SCID and M-PTSD was similar in the standard (84%) and field (84%) conditions and both were similar to previously reported agreement rates of about 85% to 86% (Engdahl et al., 1996; Kulka et al., 1991; Page, 1992).

Both POW samples were random samples of community-residing veterans, so the similarity of the figures between the two samples in Tables 1 and 2 might have been expected. On the other hand, the assessment

Table 2. Current PTSD status as assessed by the M-PTSD and the SCID under standard and field conditions

	M-PTSD status					SCID status				
	Std.	n	Field	n	X^2	Std.	n	Field	n	X^2
POW group	28%	262	28%	262	0.35	29%	262	29%	284	<0.01
Combat group	22%	110	15%	124	4.48*	20%	110	9%	135	2.48
Total sample	27%	372	22%	386		27%	372	25%	419	

Note. PTSD = post-traumatic stress disorder. M-PTSD = Mississippi Scale for Combat-Related Post-traumatic Stress Disorder. SCID = Structured Clinical Interview for DSM-III-R, PTSD module. Std. = Minnesota sample. Field = national sample. M-PTSD cutoff score = 89.
* $p < 0.01$.

DOI: 10.1002/mpr

conditions between the two samples were striking. In contrast with the standard condition sample, diagnoses in the field sample were not necessarily based on rigorously administered structured interviews by clinicians who were adequately trained and monitored. In fact, our review of the national field sample data did locate a number of instances of inappropriate use of diagnostic decision rules that led to inaccurate diagnoses in 28 cases. This makes the uniform performances of the SCID and the M-PTSD that we observed all the more impressive.

Despite the overall similarity in performance across settings, some differences between the two samples are worth noting. As a group, the national examiners were experienced in the clinical assessment of PTSD and some were no doubt familiar with the SCID. However, the group was drawn from several professions, and often from outside the VA system. This resulted in SCIDs of mixed accuracy and completeness. As has been noted, some clear misdiagnoses occurred in the field sample interviews and there were inconsistent recording practices that led to unnecessary missing data. In contrast, such errors were not found in the Minnesota data sets, which were administered under more standardized conditions. Although the overall level of SCID and M-PTSD agreement in the national sample prior to correction (as reported in Page, 1992) did not differ substantially from the agreement found in this study after corrections were made (81% versus 84%, in the original and corrected data sets, respectively), the level of errors found suggest that maintaining standards for training and administration on structured interviews can provide more complete and accurate data in future studies.

Limitations

This study compared structured clinical interviews administered by two groups of trained health professionals. As such it did not address a separate question in the literature: the validity of structured interviews administered by lay examiners versus those administered by trained clinicians (Anthony et al., 1985; Spitzer, 1993). Similarly, this study could not establish the validity of either the SCID or M-PTSD under standard or research conditions. The use of the M-PTSD as the standard for assigning a PTSD diagnosis with which to compare SCID diagnoses prevents this, as brief self-report measures cannot be used for individual clinical diagnosis. However, the use of the M-PTSD *did* allow us to compare relative rates of agreement between the SCID and this established self-report measure under differing assessment conditions, and thus permitted us to demonstrate comparable levels of agreement. The comparability of results across assessment conditions in these research settings cannot be used to directly argue for a comparability of performance in clinical settings. As Weiss (2004) suggests, in epidemiological research false positives and negatives may 'cancel each other out' with little harm to a study overall, whereas a false positive or negative diagnosis in a clinical setting can have lasting negative impacts on a patient.

Symptoms of PTSD in both samples were stable and had been experienced for decades by most subjects (Gold et al., 2000; Engdahl and Fairbank, 2001). More variability in the assessment performance of the SCID PTSD module would be expected for the first year or more after war trauma exposure when symptoms are mixed and their levels fluctuate (Port et al., 2001; Port et al., 2002), taxing the skills of the interviewer and the limits of the instrumentation. The generalizability of our observations may also be limited, as our participants were aged male combat veterans and predominately white. It is unclear to what extent results would generalize to participants from other demographic groups or to American or British service members returning from the current war in Iraq and Afghanistan. Nonetheless, our results suggest that, even under less than optimal field conditions, there is a basis for confidence in the validity of results from instruments developed under more tightly controlled conditions.

Acknowledgements

This research was supported in part by the Medical Research Service of the Department of Veterans Affairs, and by the Medical Follow-up Agency of the National Academy of Sciences.

The opinions and assertions contained herein are those of the authors and are not to be construed as reflecting the views or positions of the National Academy of Sciences, The Institute of Medicine, or the National Research Council.

References

Anthony JC, Folstein M, Romanski AJ, Vaon Korff MR, Nestadt GR, Chahal R. Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis: experience in eastern Baltimore. Arch Gen Psychiatry 1985; 42: 667–75.

DOI: 10.1002/mpr

Aziz MA, Kenford S. Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. J Psychiatr Prac 2004; 10: 307–13.

Engdahl B, Dikel T, Eberly RE, Blank A. Posttraumatic stress disorder in a community sample of former prisoners of war: a normative response to severe trauma. Am J Psychiatry 1997; 154: 1576–81.

Engdahl B, Dikel T, Eberly RE, Blank A. Comorbidity and course of psychiatric disorders in a community sample of former prisoners of war. Am J Psychiatry 1998; 155: 1740–5.

Engdahl B, Eberly RE, Blake JD. The assessment of posttraumatic stress disorder in World War II veterans. Psychol Assess 1996; 8: 445–9.

Engdahl B, Fairbank J. Former prisoners of war: highlights of empirical research. In E Gerrity, T Keane, F Tuma (eds) Mental Health Consequences of Torture and Other Severe Trauma. New York: Plenum Press, 2001.

Foa EB, Tolin DF. Comparison of the PTSD Symptom Scale – Interview Version to the Clinician Administered PTSD Scale. J Trauma Stress 2000; 13: 181–92.

Gold P, Engdahl B, Frueh C, Blake R. Trauma exposure, resilience, social support, and PTSD construct validity among former POWS. Soc Psychiatry Psychiatr Epidemiol 2000; 35: 36–42.

Keane TM, Caddell JM, Taylor KL. Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: three studies in reliability. J Consult Clin Psychol 1988; 56: 85–90.

Kulka RA, Schlenger WE, Fairbank JA, Jordan BK, Hough RL, Marmar CR, Weiss DS. Assessment of posttraumatic stress disorder in the community: prospects and pitfalls from recent studies of Vietnam veterans. Psychol Assess 1991; 3: 547–60.

Page WF. Using longitudinal data to estimate nonresponse bias. Soc Psychiatry Psychiatr Epidemiol 1991; 26: 127–31.

Page WF. The Health of Former Prisoners of War: Results from the Medical Examination Survey of former POWs of World War II and the Korean Conflict. Washington DC: National Academy Press, 1992.

Port C, Engdahl B, Frazier P. A longitudinal and retrospective study of PTSD among older prisoners of war. Am J Psychiatry 2001; 158: 1474–9.

Port C, Engdahl B, Frazier P, Eberly R. Factors related to the long-term course of PTSD in older ex-prisoners of war. J Clin Geropsychol 2002; 8: 203–14.

Schlenger WE, Allison M. Structured Clinical Interview for DSM-III-R-non-patient edition (SCID-NP-V: Supplemental User's Guide). Research Triangle Park NC: National Vietnam Veterans Readjustment Study, 1987.

Schlenger WE, Jordan BK, Caddell JM, Ebert L, Fairbank JA. Epidemiological Methods for Assessing Trauma and PTSD. In JP Wilson, TM Keane (eds) Assessing Psychological Trauma and PTSD. New York: Guilford Press, 2004.

Spitzer RL, Williams JB. Structured Clinical Interview for DSM-III-R Non-patient Edition (Modified for Vietnam Veterans Readjustment Study 4/1/87). New York: New York State Psychiatric Institute, Biometrics Research, 1986.

Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? Compr Psychiatry 1993; 24: 399–411.

Weathers FW, Keane TM, Davidson J. Clinician-Administered PTSD Scale: a review of the first ten years of research. Depress Anxiety 2001; 13: 132–56.

Weiss DS (2004). Structured clinical interview techniques for PTSD. In JP Wilson, TM Keane (eds) Assessing Psychological Trauma and PTSD. New York: Guilford Press, 2004.

Correspondence: Brian Engdahl, Psychology Service (116B), VA Medical Center, One Veterans Drive, Minneapolis, MN 55417.
Telephone: (+1) 612 725-2073.
Fax: (+1) 612 727-5964.
Email: brian.engdahl@med.va.gov

DOI: 10.1002/mpr