

Impact of Evidence-Based Standardized Assessment on the Disability Clinical Interview for Diagnosis of Service-Connected PTSD: A Cluster-Randomized Trial

Theodore Speroff,^{1,2} Patricia L. Sinnott,^{3,4} Brian Marx,^{5,6} Richard R. Owen,^{7,8} James C. Jackson,^{1,2} Robert Greevy,^{1,2} Nina Sayer,^{9,10} Maureen Murdoch,^{9,10} Andrea C. Shane,³ Jeffrey Smith,^{8,11} JoAnn Alvarez,^{1,2} Samuel K. Nwosu,^{1,2} Terence Keane,^{5,6} Frank Weathers,¹² Paula P. Schnurr,^{13,14} and Matthew J. Friedman^{13,14}

¹Center for Health Services Research, Geriatric Research, Education, and Clinical Center (GRECC), Veterans Affairs Tennessee Valley Healthcare System, Nashville, Tennessee, USA

²Vanderbilt University School of Medicine, Nashville, Tennessee, USA

³Health Economics Resource Center, Veterans Affairs Palo Alto Health Care System, Menlo Park, California, USA

⁴Center for Health Policy/Center for Primary Care and Outcomes Research, Stanford University, Stanford, California, USA

⁵Veterans Affairs National Center for Posttraumatic Stress Disorder, Veterans Affairs Boston Healthcare System, Boston, Massachusetts, USA

⁶Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts, USA

⁷Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, USA

⁸Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

⁹Center for Chronic Disease Outcomes Research, Minneapolis Veterans Affairs Medical Center, Minneapolis, Minnesota, USA

¹⁰University of Minnesota School of Medicine, Minneapolis, Minnesota, USA

¹¹Veterans Affairs Mental Health Quality Enhancement Research Initiative (MH QUERI), Central Arkansas Veterans Affairs Healthcare System, North Little Rock, Arkansas, USA

¹²Department of Psychology, Auburn University, Auburn, Alabama, USA

¹³Veterans Affairs National Center for Posttraumatic Stress Disorder, White River Junction Veterans Affairs Medical Center, White River Junction, Vermont, USA

¹⁴Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

Posttraumatic stress disorder (PTSD) is one of the fastest growing compensated medical conditions. The present study compared usual disability examiner practices for PTSD with a standardized assessment that incorporates evidence-based assessments. The design was a multicenter, cluster randomized, parallel-group study involving 33 clinical examiners and 384 veterans at 6 Veterans Affairs medical centers. The standardized group incorporated the Clinician Administered PTSD Scale and the World Health Organization Disability Assessment Schedule-II into their assessment interview. The main outcome measures were completeness and accuracy of PTSD diagnosis and completeness of functional assessment. The standardized assessments were 85% complete for diagnosis compared to 30% for nonstandardized assessments ($p < .001$), and, for functional impairment, 76% versus 3% ($p < .001$). The findings demonstrate that the quality of PTSD disability examination would be improved by using evidence-based assessment.

This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development QUERI program. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. This project is supported by the VA Health Service Research and Development QUERI Program, SDR 06–331. (Clinicaltrials.gov registration number NCT00699660)

Correspondence concerning this article should be addressed to Ted Speroff, Department of Medicine, Center for Health Services Research, 6000 Medical Center East, Vanderbilt University School of Medicine, Nashville, TN 37232. E-mail: ted.speroff@vanderbilt.edu

Published 2012. This article is a US Government work and is in the public domain in the USA. View this article online at wileyonlinelibrary.com
DOI: 10.1002/jts.21759

Posttraumatic stress disorder (PTSD) is an anxiety disorder that may develop in response to high magnitude stressors such as natural disasters, serious accidents, critical medical conditions, sexual assault, violence, war, and terrorism (American Psychiatric Association [APA], 2000; Institute of Medicine, 2006; World Health Organization [WHO], 2004). PTSD symptoms include intrusive memories, nightmares, and flashbacks of the traumatic stressors; avoidance of reminders, emotional numbing, and detachment from others; and hyperarousal, including startle reactions, hypervigilance, difficulty with sleep, concentration problems, and anger (APA, 2000). Symptoms can be enduring and patients with unrecognized PTSD are often treated in clinical practice for a variety of other mental and physical health problems (Keane, Weathers, & Foa, 2000).

PTSD is a growing burden to public health with considerable economic cost to society (Bilmes, 2007; Kessler, 2000; Savoca & Rosenheck, 2000; WHO, 2004). Many patients with PTSD experience persistent impairment in work or school performance, marital and family functioning, interpersonal relationships, and social and community activities (Kennedy, 2002; Kessler, 2000; Schnurr, Lunney, Bovin, & Marx, 2009; Smith, Schnurr, & Rosenheck, 2005). Subsequently, afflicted patients often seek compensation from Social Security or the U.S. Department of Veterans Affairs (VA).

PTSD has been among the fastest growing compensated conditions for both Social Security and the VA (Kennedy, 2002; U.S. Department of Veteran Affairs VA Office of Inspector General, 2006). The VA's disability compensation program is second only to Social Security Disability Insurance in size and currently covers almost 3 million veterans (U.S. Department of Veteran Affairs Veterans Benefits Administration, 2010). Between 1999 and 2004, benefit payments for service-connected PTSD have increased 149% (up to over \$4 billion annually) compared to just 42% for all other service-related disabilities (U.S. Department of Veteran Affairs VA Office of Inspector General, 2006). Almost 400,000 veterans received compensation for PTSD during 2010, a 222% increase from 1999 (U.S. Department of Veteran Affairs Veterans Benefits Administration, 2010). These rapidly rising costs of PTSD compensation prompted a government investigation that found wide regional variations in the rates of service-connected PTSD and attributed this variance in part to variation in the diagnostic examination (U.S. Department of Veteran Affairs VA Office of Inspector General, 2006).

The examination request for service-connected PTSD specifies the need for a diagnosis according to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV-TR) (American Psychiatric Association, 2000) supported by the findings on the examination and requires the examiner to describe changes in psychosocial functional status and quality of life. It is incumbent upon the examiner to administer a complete clinical interview and provide a thorough disability report. Most clinicians in the VA use unstructured clinical interviews when conducting these PTSD disability examinations (Jackson et al., 2011). In an early study, Steiner, Tebes, Sledge,

& Walker (1995) found substantial differences in determination of mental disorder between unstructured and structured interviews but concluded that although structured clinical interviews were more orderly, it was not clear which is more comprehensive. Additional studies have indicated that unstructured clinical interviews may lack thoroughness, produce insufficient data collection, and result in incorrect application of diagnostic criteria (Shear et al., 2000; Skodol, Williams, Spitzer, Gibbon, & Kass, 1984). Furthermore, given that structured, standardized interviews are used routinely to enhance the rigor of research studies on PTSD, we speculated that incorporating these measures into the disability examination might produce greater consistency and validity compared to unstructured interviews.

A surprising lack of studies exists on assessment of PTSD for compensation. In fact, there have been no published we could locate on the degree to which inclusion of evidence-based, standardized disability assessment methods would increase completeness of diagnostic information, decrease variation, and improve diagnostic accuracy for any condition in any disability system. We addressed this evidence gap by conducting a cluster randomized controlled trial, comparing usual practice, non-standardized clinical interviews in a sample of veterans seeking PTSD disability compensation from the VA with standardized assessments that incorporated well-validated, evidence-based assessments for functional impairment and PTSD diagnosis. We hypothesized that veterans in the standardized assessment condition would receive more complete and accurate assessment of the *DSM-IV-TR* diagnostic components of PTSD and related functional impairment than veterans in the usual practice condition.

Method

Participants and Procedure

Veterans Affairs medical centers with high annual volume of PTSD examinations where nonstandardized disability assessment was usual practice were eligible for the study. Six medical centers with commitment from senior management participated in the study. Clinical examiners were psychologists and psychiatrists who provided written informed consent. Participants were veterans who had submitted a claim for PTSD disability and were referred to the medical center by the Veterans Benefits Administration (VBA) for an initial PTSD disability assessment. Veterans were excluded if they were under 18 years of age, cognitively impaired, or refused informed consent. Institutional review board approval was obtained from each medical center.

This was a multicenter, cluster randomized, parallel-group study comparing standardized with nonstandardized initial disability assessment of PTSD within the VA. Statistical randomization with equal allocation within each medical center was at the clinician level to minimize potential contamination of interview techniques at the participant level. Veterans were assigned to clinicians for a PTSD assessment based on next available

appointment by schedulers blind to the group assignment of the clinical examiner.

Group interventions. Examiners in the nonstandardized disability assessment group continued to use their usual practice for clinical interviews, use of adjunct psychological testing was permitted, but such testing was not required by our study protocol. The clinical examiners in the standardized disability assessment group integrated the Clinician-Administered PTSD Scale (CAPS) to diagnose PTSD and the World Health Organization Disability Assessment Schedule II (WHODAS-II) to assess for functional impairment into their clinical interview (Weathers et al., 2004; WHO, 2000). Clinicians in the standardized disability assessment condition underwent 10 hours of formal CAPS and WHODAS-II training that included role playing. The first two study examinations following this training were reviewed and additional feedback was provided to confirm and refine their assessment techniques. All examiners completed the VA certification training for compensation and pension examinations that reviews diagnosis of PTSD and provides instruction for a thorough interview and a complete PTSD disability examination report.

The CAPS is a structured diagnostic interview that requires the interviewer to rate the frequency and intensity of *DSM-IV* symptoms of PTSD (Weathers et al., 2004). Standardized administration is accomplished through carefully worded prompts and scale anchors with explicit behavioral referents (Weathers et al., 2004; Weathers, Keane, & Foa, 2009). Initial prompt questions target each core symptom, and follow-up prompts help clinicians clarify the inquiry as needed for linkage between symptom and trauma. The CAPS has been extensively validated and is the most widely used evidence-based PTSD interview (Watson et al., 2005; Weathers et al., 2004, 2009; Weathers, Keane, & Davidson, 2001). Similarly, the WHODAS-II has been widely used to measure functional status across a wide-range of disabilities (Chwastiak & Von, 2003; Kennedy, 2002; World Health Organization, 2000). The 36-item WHODAS-II is administered through a standardized clinical interview with probing to comprehensively assess functional impairment in communication, mobility, self-care, getting along with others, daily activities, work, and participation in society. The objective of the CAPS and WHODAS-II is to semistructure the clinical interview, allowing for timely appropriate in-depth probes, which then produces standardization for measuring the defining criteria for diagnosis of PTSD and its impact on psychosocial functioning.

Study procedures. Immediately following confirmation that the examination had been scheduled, veterans were sent a letter about the study and a research assistant telephoned the veteran to determine interest in participation. If interested, the research assistant made arrangements to greet the veteran at the scheduled examination to evaluate exclusion criteria, obtain written informed consent, and initiate data collection. Following consent, the research assistant accompanied the veteran to

the room where the clinical examiner conducted and digitally recorded the examination. Upon conclusion of the interview, the veteran completed self-report study questionnaires on satisfaction and functional impairment (not part of the disability examination and extraneous to this analysis) and received \$10.00 in reimbursement. All disability examinations occurred in the usual course of conducting the exam and submitting the report to VBA for disability rating within a time window of 30 calendar days.

Digital audio recordings were electronically transferred to the VA National Center for PTSD (NC-PTSD) for blinded review. Clinical psychologists at the NC-PTSD (expert reviewers) listened to the recordings and scored the clinical interview for diagnostic and functional assessment. All 10 expert reviewers had at least 5 years of clinical experience and seven had at least 15 years of experience assessing and treating veterans with PTSD outside of research protocols. Pilot testing was conducted to refine the measurement instruments. For scoring the audio recordings, pilot cases were reviewed until the expert reviewers obtained a kappa coefficient of .89. During the study, exams received only single reviews to keep within budget and timelines.

Measures

The study outcome measures were completeness and accuracy of the clinical interview.

Overall completeness of information. NC-PTSD expert reviewers classified each case as “yes” or “no” on whether all relevant and important information regarding diagnostic and functional status was included in the clinical interview.

Completeness of the diagnostic interview. The NC-PTSD expert reviewers used the PTSD module from the Structured Clinical Interview for the *DSM-IV* (SCID PTSD) to independently evaluate whether the clinical interview was sufficiently complete in addressing all the diagnostic criteria for PTSD. The SCID PTSD was slightly modified, such that the response options for each of the diagnostic items included absent, subthreshold, threshold, and unable to determine. The unable to determine option was included to reflect missing and incomplete diagnostic information and to allow evaluation of the comprehensiveness of the examination. The completeness of the diagnostic interview was the proportion of 21 questions on the SCID PTSD form that were not marked “unable to determine” during review of the recording. Completeness scores, also calculated for each *DSM-IV* criterion, ranged between 0% for an incomplete clinical interview and 100% for a fully complete interview.

Thoroughness of the assessment. The NC-PTSD expert rated each functional domain for thoroughness of assessment along a 5-point scale ranging from 1 = *not assessed at all* to 5 = *thoroughly assessed* and rescaled to range from 0% to 100%.

The ratings were averaged to yield an overall thoroughness score.

Diagnostic impression. The NC-PTSD reviewers provided summary scores for *DSM-IV* Criteria A (trauma exposure), B (intrusion), C (avoidance and numbing), D (elevated arousal), E (duration), and F (functional impairment) and categorized the case into one of three diagnostic impressions: PTSD present, PTSD absent, and Unable to Determine reflecting incomplete evidence in the audio of the clinical interview to either rule in or rule out PTSD. The NC-PTSD reviewers rated the unable to determine classification from almost certain that PTSD is not present (−3) to almost certain PTSD is present (+3).

Concordance. The clinical examiners were required to provide a PTSD diagnosis in their report; these diagnoses were extracted and classified into the diagnostic categories of PTSD present versus PTSD absent. The clinical examiner diagnoses were compared with the NC-PTSD expert reviewers' diagnoses as the reference standard.

Our sample size calculation yielded 466 for a power of .80 to detect a 10% absolute difference in sensitivity and adjusted for intraclass correlation.

Data Analysis

Description of the study population and main results are presented as proportions (%), medians (*Mdn*), interquartile ranges (IQR), means (*M*), and standard deviations (*SD*) as appropriate. Estimates and *p* values of tests for statistical significance were derived from linear and logistic mixed effects regression with clinical examiner stratified by study group as a random effect and study group as a fixed effect. Covariate adjusted models for continuous outcomes included the prespecified years of experience doing PTSD disability examinations, use of an examination report template, use of psychological tests, time spent reviewing records, veteran age, veteran education, study site, and expert reviewer. Although a randomized study design will yield unbiased estimates of the group effect regardless of adjustment for examiner and patient level characteristics, adjusting for prespecified covariates may increase the precision of the estimates (Duflo, Glennerster, & Kremer, 2007). The number of covariates in regression models was limited to ensure the effective sample size remained 10 times greater than the degrees of freedom in the model. A variance components model was fit with random effects to estimate the proportion of variance between examiners and variance attributed to differences between sites. Statistical significance was interpreted at a .05 level; analyses were computed using R (R Development Core Team, 2006).

Results

The numbers for study retention with reasons for exclusion are shown in Figure 1. Although six clinical examiners dropped out

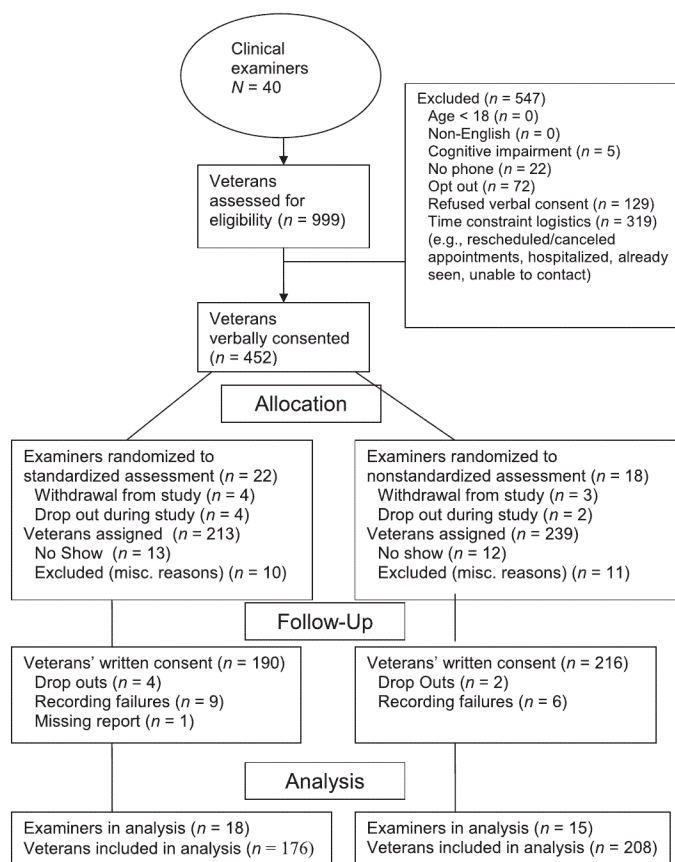


Figure 1. Cluster randomized trial recruitment and retention flow diagram.

due to promotions and job mobility, their data were retained, yielding 33 examiners in analysis. Between March 17, 2009 and September 29, 2010, 999 eligible veterans were assigned to the participating clinicians. Exclusions were due mainly to time constraint logistics related to the 3-week window for study enrollment. We obtained written informed consent for 406 veterans (41%) and obtained data from 384 (95%). Table 1 lists participant demographic characteristics by condition.

Table 2 shows that the NC-PTSD experts rated 0.5% of the nonstandardized exams as complete in diagnostic and functional information compared with 62% of the standardized exams, $\chi^2(1, N = 384) = 176.13, p < .001$. The average score for completeness of assessing the *DSM-IV* criteria for PTSD was 85% for the standardized disability assessment condition and 30% for the nonstandardized disability assessment condition, $t(24) = 11.42, p < .001$ (see Table 2); completeness was uniformly greater among examiners in the standardized disability assessment condition across the core diagnostic criteria. In addition, there was greater consistency in diagnostic assessment between medical centers and clinical examiners within the standardized condition. The completeness score standard deviation among examiners was .002 for the standardized disability assessment group and 17.36 for the nonstandardized disability assessment group (likelihood ratio = 7.16, $p = .047$). In the

Table 1
Characteristics of Veterans and Clinicians by Disability Assessment Group

Variable	Standardized		Nonstandardized	
	<i>n</i>	% or IQR	<i>n</i>	% or IQR
Veterans ^a				
Female	6	3	13	6
Median age	60	41–63	59	41–62
Age 18–50 years	57	32	81	39
Age >50 years	119	68	127	61
Marital status				
Married	108	61	130	62
Single	28	16	35	17
Separated	32	18	37	18
Other/missing	8	5	5	2
Education				
<High school	9	5	13	6
High school or GED	68	39	64	31
Vocational school	9	5	16	8
College	90	51	115	56
Race/ethnicity				
African American	44	25	56	27
Caucasian	103	59	124	60
Latino	16	9	15	7
Other	13	8	13	5
Active duty combat	162	92	187	90
Era ^b				
OEF/OIF	55	31	81	39
Desert Shield/Gulf war	19	11	37	18
Vietnam	106	60	114	55
Korea	4	2	5	2
WWII	2	1	3	1
Other	4	2	6	3
Branch of service ^b				
Army	117	66	151	73
Marine	27	15	36	17
Navy	19	11	14	7
Air Force	16	9	15	7
Clinical examiners				
Female	13	72	8	53
Race/Ethnicity				
African American	0	0	0	0
Caucasian	15	83	14	93
Latino	2	11	0	0
Other	1	6	1	7
Psychologists	17	94	15	100
Median years' experience	2.0	0.6–7.3	3.0	1–8

Note. IQR = Interquartile ranges; OEF/OIF = Operation Enduring Freedom/Operation Iraqi Freedom.

^a*N* = 176 veterans in the standardized condition and *N* = 208 in the nonstandardized group and *N* = 18 for clinical examiners in the standardized and *N* = 15 in the nonstandardized group. ^bSum is greater than sample size because some veterans were in multiple war eras and branches of service.

nonstandardized disability assessment condition, medical centers and clinical examiners within medical centers contributed 10% and 17%, respectively, to variation whereas nearly 100% of the variation in the standardized disability assessment condition was between cases.

The thoroughness of the functional assessment (Table 2) was greater in the standardized exam compared with the nonstandardized exam (76% vs. 3%, respectively), $t(24) = 37.89$, $p < .001$. The completeness of the functional assessment in the nonstandardized disability assessment condition was very low in each domain due to insufficient information.

The concurrence between the NC-PTSD expert and the examiner diagnoses is shown in Table 3. Agreement of PTSD diagnosis was 93% for the standardized exam group and 77% for the nonstandardized exam group. The caveat, however, is that the NC-PTSD experts found that 64% of the cases could not be determined based on the lack of completeness of the interview. The percentage of inability to determine PTSD diagnosis was higher in nonstandardized examinations (88%, 183 of 208 cases) than in standardized examinations (36%, 63 of 176 cases), $\chi^2(1, N = 384) = 112.77$, $p < .01$.

The source of this diagnostic uncertainty was uniformly above 50% for the nonstandardized exams (Table 4) due to failure to ascertain whether symptoms were linked to the trauma event, a critical component of the PTSD *DSM-IV* diagnosis. The NC-PTSD reviewers were asked to rate the likelihood of PTSD based on available data along a scale spanning from -3 (likely that PTSD is absent) to $+3$ (likely that PTSD is present). Ratings of $+3$ and $+2$ were then attributed to the expert's diagnosis of PTSD and ratings of -3 and -2 were assimilated into the absence of PTSD. The cases with ratings in the middle ($+1$, 0 , and -1) remained in the region of absolute uncertainty of PTSD diagnosis.

Table 5 shows the concurrence between the NC-PTSD expert and examiner diagnoses using the resolution of unable to determine classification. The proportion of exams in which the reviewer was uncertain was 10% (18 of 176 cases) versus 52% (109 of 208 cases) in the standardized and nonstandardized condition, respectively, $\chi^2(1, N = 384) = 76.61$, $p < .001$. Agreement of PTSD diagnosis was 92% for standardized disability assessment and 78% for nonstandardized disability assessment, yielding an unadjusted difference of 14%, 95% confidence interval (CI) = [3.6, 24.1], favoring the standardized exam and adjusted difference of 11.3%, $t(31) = 1.67$, $p = .10$. Logistic regression controlling for examiner as a random effect yielded an odds ratio (OR) of 2.0, 95% CI = [0.7, 6.2]. The positive predictive value (PPV) was 87% (109 of 125 cases) in the standardized condition and 53% (66 of 125 cases) in the nonstandardized condition, yielding an unadjusted difference in PPV of 34%, 95% CI = [23, 44], favoring the standardized disability exam, adjusted difference of 35% (95% CI = [23, 47]), and the logistic regression including examiner as a random effect yielded an OR of 6.3, 95% CI = [3.2, 12.4].

In the standardized disability assessment group, the prevalence of PTSD according to expert reviewer diagnoses was 68%

Table 2

Differences by Disability Assessment Group on Measures of Completeness of Diagnosis and Functioning

Variable	Standardized		Nonstandardized		χ^2 or Wald <i>t</i>
	<i>M</i> or <i>n</i>	<i>SD</i> or %	<i>M</i> or <i>n</i>	<i>SD</i> or %	
All diagnostic and functional status information obtained	109	62	1	0.5	176.13***
Diagnostic assessment completeness					
Total	84.8	17.3	29.8	24.4	11.42***
Criterion A: Exposure	84.1	26.2	67.7	32.2	0.90
Criterion B: Reexperiencing	85.6	22.5	25.6	30.2	7.84***
Criterion C: Avoidance/numbing	82.9	22.5	21.6	29.3	8.00***
Criterion D: Hyperarousal	84.7	24.3	27.5	30.5	8.96***
Criterion E: Duration	166	94	89	43	4.77***
Criterion F: Impairment	154	88	64	31	5.92***
Functional assessment completeness					
Thoroughness	76.2	24.4	3.1	7.5	37.89***
Communication domain	81.1	26.6	0.7	4.9	37.32***
Mobility domain	81.4	26.0	1.7	7.6	34.60***
Self-care domain	79.0	27.9	2.5	9.3	26.63***
Getting along domain	80.7	25.6	6.0	14.1	27.20***
Activities domain	81.1	25.8	2.4	10.1	32.85***
Work domain	49.6	42.8	5.9	13.2	12.32***
Social domain	80.4	26.8	2.2	8.2	33.59***

Note. $\chi^2 df = 1$. The Wald test generates *t* values with 24 degrees of freedom using the estimated standard error for the mixed effects regression coefficient. Adjusted analysis had *N* = 382 veterans and 33 examiners due to missing values on examiner preparation time for two cases. Scores for total completeness and diagnosis Criteria A–D range between 0% = fully incomplete to 100% = fully complete. Scores for thoroughness of assessment and the domain scores range from 0% to 100%.

*** *p* < .001.

(119 of 176 cases), which is similar to the 71% prevalence (125 of 176 cases) by the clinical examiners.

Discussion

The increase in prevalence of PTSD, high cost of disability payment, variability in disability determination, and potential insufficient evaluation have raised concerns about the manner in which the disability examination for PTSD is conducted (Andreasen, 2010; Institute of Medicine, 2006, 2007; U.S. Department of Veteran Affairs VA Office of Inspector General, 2006). In a survey of VA clinical examiners, we found that the CAPS or other structured interview methods are rarely used in disability assessment (Jackson et al., 2011). The more routine and common practice is the open-ended, unstructured clinical interview concurrent with the use of a report-writing template. The findings of this study show that administering a standardized disability assessment resulted in more complete coverage of functional impairment and PTSD symptoms. Standardized assessment elicited an increase in relevant information and nearly eliminated variation between examiners and medical centers. Furthermore, this study found that standardized assessment substantially diminished the uncertainty in diagnosis, and increased concordance of diagnosis with the NC-PTSD experts.

The standardized exam was more sensitive than routine examination, but did not result in a significant change in the overall prevalence of diagnosed PTSD.

In our study, the usual practice, nonstandardized clinical interview, produced incomplete elicitation of diagnostic information. This result complements the literature reporting that standardized psychiatric interviews in the clinical setting are more accurate, thorough, and reliable than unstructured psychiatric interviews (Miller, 2001; Miller, Dasher, Collins, Griffiths, & Brown, 2001; Shear et al., 2000; Sheehan et al., 1998; Skodol, Williams, Spitzer, Gibbon, & Kass, 1984). Miller et al. (2001) found that compared with a consensus reference standard, the sensitivity of nonstandardized clinical interviews for psychiatric diagnoses was 54% and the sensitivity of the SCID-CV was 86%. Other studies have found that diagnoses formulated by using the standardized SCID-CV had very low agreement with diagnoses using a nonstandardized clinical interview (Shear et al., 2000; Ventura, Liberman, Green, Shaner, & Mintz, 1998). Similar to our findings, the difference in accuracy was attributed to the comprehensiveness of the evaluation of diagnostic criteria; in this prior research, clinicians using nonstandardized interviews evaluated only 53% of the key criteria whereas clinicians using structured interviews evaluated 100% of the key criteria (Miller et al., 2001). Our study adds substantially to this literature and reinforces the importance of

Table 3

Concurrence of Clinical Examiner Diagnosis With Expert Reviewer by Disability Assessment Group

Variable	Standardized							Nonstandardized						
	PTSD present		UTD ^a		PTSD absent		Total	PTSD present		UTD		PTSD absent		Total
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>
Clinical examiner diagnosis														
PTSD present	69	93	49	78	7	18	125	13	77	110	60	2	25	125
PTSD absent	5	7	14	22	32	82	51	4	24	73	40	6	75	83
Total	74		63		39		176	17		183		8		208

Note. PTSD = posttraumatic stress disorder; UTD = unable to determine.

^aNC-PTSD expert reviewers classified diagnoses as PTSD present, PTSD absent, or UTD.

systematic, standardized assessment of mental health conditions to obtain diagnostic consistency across clinicians in real-world settings (Antony & Rowa, 2005; Hunsley & Mash, 2005, 2007). The importance of diagnostic accuracy and validity is of particular importance in the medicolegal context of a disability examination where the diagnosis has clear implications for equity in distribution of financial compensation and priority access to VA health care. The public policy benefit of improved disability assessment is ensuring that compensation is provided to those individuals who meet disability criteria.

To meet diagnostic criteria for PTSD, an individual must not only endorse the requisite number of PTSD symptoms, but also report that these symptoms have resulted in clinically significant distress or impairment in social, occupational, or other important areas of functioning. It is easy to see how many PTSD symptoms can lead to difficulties, for example, symptoms of numbing could easily lead to relationship difficulties, and it is easy to imagine how angry outbursts could cause trouble at work. Practicing clinicians, as those in the nonstandardized in-

terview group, are more likely to assess only overall level of functioning, whereas a more detailed, thorough functional assessment may provide context to the PTSD diagnosis and guide treatment interventions.

There are several limitations to this study. Sources of information external to the clinical interview such as psychological testing or charts were not available to the NC-PTSD expert reviewers. The utilization of psychological testing, however, was low; the time spent by examiners conducting chart review was approximately 27 minutes on average. Furthermore, our analysis adjusting for time spent in chart review and presence of testing did not alter results. Another limitation is that veterans in our study did not undergo a concurrent, independent evaluation for assessment of PTSD. The focus of our study comparing standardized with nonstandardized assessment was on the impact on the clinical interview rather than the efficacy of the CAPS and WHODAS-II, which already have a solid base of psychometric evidence. Another limitation is that despite study procedures for blinding the expert reviewers, NC-PTSD experts

Table 4

Differences by Assessment Group on DSM-IV PTSD Criteria Ratings Based on Expert Review Using Recorded Interview

Criterion	Standardized										Nonstandardized										χ^2
	Present		UTD		Absent		Not ask		No probe		Present		UTD		Absent		Not ask		No probe		
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Exposure	116	65	51	29	10	6	35	20	16	9	90	43	115	55	3	1	81	39	34	16	27.77
Intrusion	151	86	51	11	6	14	8	4	3	2	77	37	128	62	3	1	105	50	23	12	126.18
Avoidance																					
Numbing	112	64	35	20	29	16	5	3	30	17	32	15	171	82	5	2	140	67	31	15	149.92
Hyperarousal	136	77	19	11	21	12	5	3	14	8	60	29	144	69	4	2	99	48	45	22	133.25
Duration	162	92	10	6	4	2	8	4	2	2	88	42	119	57	1	0	83	40	36	17	113.47
Impairment	132	75	22	12	22	12	6	3	16	9	59	28	144	69	5	2	74	35	70	34	125.02

Note. DSM-IV = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.); PTSD = posttraumatic stress disorders; UTD = unable to determine. Criteria A-F were classified as present, absent, or UTD, which was subcategorized further as failure to ask or lack of probing. The χ^2 has 1 degree of freedom and tests for differences in the proportion of UTD; all $p < .001$.

Table 5

Differences by Assessment Group on PTSD Criteria Using Narrower Criteria for UTD

Variable	NC-PTSD expert reviewer diagnostic classification ^a													
	Standardized disability assessment							Nonstandardized disability assessment						
	PTSD present		Un-certain		PTSD absent		Total	PTSD present		Un-certain		PTSD absent		Total
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>
Clinical examiner diagnosis														
PTSD present	109	92	9	50	7	18	125	66	78	57	52	2	14	125
PTSD absent	10	8	9	50	32	82	51	19	22	52	48	12	86	83
Total	119		18		39		176	85		109		14		208

Note. ^aNC-PTSD expert reviewers classified diagnoses as PTSD present, PTSD absent, or Uncertain. The Unable to Determine classifications were rated from -3 (PTSD absent) to +3 (PTSD present) and ratings of -1, 0, +1 were retained as "Uncertain".

may have been able to guess group assignment. We attempted to minimize any potential bias by ensuring that reviewers received intensive training to assure interrater consistency, and by using SCID-PTSD symptom scoring that required very clear and explicit information about occurrence of symptoms in the clinical interviews. In addition, the two study groups of clinicians could have differed in characteristics and experience that we did not measure. All VA clinicians who administer disability examinations for PTSD, however, must obtain course certification on PTSD disability examinations and the elements required in a disability report, providing greater assurance of equivalent baseline knowledge and skills in the two study groups. Finally, participants may have differed from nonparticipants. Although application of privacy rules prevented us from collecting personal data from nonparticipants to rule out differences, logistical constraints of the narrow study time window for recruitment rather than sampling differences accounted for the majority of nonparticipation. Despite these limitations, our study had numerous strengths. We conducted the study within the real-world of disability examination; our effect size was very large for testing group differences in interview completeness and sufficiently large to detect significance in diagnostic PPV, multiple medical centers participated, and blinded scheduling of veterans to clinical examiners produced equivalence between study groups.

In conclusion, use of the CAPS and WHODAS-II produced a more comprehensive standardized disability assessment than did the usual practice, nonstandardized clinical interview. The standardized clinical interview reduced variation between clinicians and sites, providing greater assurance that claims are treated equitably in clinical assessment. This study has important implications not only for PTSD disability assessment, but also for assessment of PTSD more broadly and for the role of evidence-based assessment of psychiatric disorders in the actual field conditions of the clinical setting (Barlow, 2005; Erbes, Dikel, Eberly, Page, & Engdahl, 2006; Zimmerman, 2003). Our study indicates that evidence-based, standardized disability assessment for PTSD would enhance the clinician's determination of a PTSD diagnosis and functional impairment

and make the disability examination process more reliable and accountable.

References

- Andreasen, N. C. (2010). Posttraumatic stress disorder: a history and critique. *Annals New York Academy of Science*, 1208, 67–71. doi:10.1111/j.1749-6632.2010.05699.x
- Antony, M. M., & Rowa, K. (2005). Evidence-based assessment of anxiety disorders in adults. *Psychological Assessment*, 17, 256–266. doi:10.1037/1040-3590.17.3.256
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Barlow, D. H. (2005). What's new about evidence-based assessment? *Psychological Assessment*, 17, 308–311. doi:10.1037/1040-3590.17.3.308
- Bilmes, L. (2007). *Soldiers returning from Iraq and Afghanistan: The long-term costs of providing veterans medical care and disability benefits* (Faculty Research Working Papers Series RWP07-001; pp.1–20). Cambridge, MA: John F. Kennedy School of Government, Harvard University. doi:10.2139/ssrn.939657
- Chwastiak, L. A., & Von, K. M. (2003). Disability in depression and back pain: Evaluation of the World Health Organization Disability Assessment Schedule (WHODAS II) in a primary care setting. *Journal of Clinical Epidemiology*, 56, 507–514.
- Duflo E., Glennerster R., & Kremer, M. (2007). *Using randomization in development economics: A toolkit* (NBER Technical Working Paper No. 333). Cambridge, MA: National Bureau of Economic Research.
- Erbes, C. R., Dikel, T. N., Eberly, R. E., Page, W. F., & Engdahl, B. E. (2006). A comparative study of posttraumatic stress disorder assessment under standard conditions and in the field. *International Journal of Methods in Psychiatric Research*, 15, 57–63. doi:10.1002/mpr.185
- Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*, 17, 251–255. doi:10.1037/1040-3590.17.3.251
- Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review Clinical Psychology*, 3, 29–51. doi:10.1146/annurev.clinpsy.3.022806.091419
- Institute of Medicine. (2006). *Posttraumatic stress disorder: Diagnosis and assessment*. Washington, DC: National Academies Press.

- Institute of Medicine. (2007). *A 21st century system for evaluating veterans for disability benefits*. Washington, DC: National Academies Press.
- Jackson, J. C., Sinnott, P. L., Marx, B. P., Murdoch, M., Sayer, N. A., Alvarez, J. M., . . . Speroff, T. (2011). Variations in practices and attitudes of clinicians assessing PTSD-related disability among veterans. *Journal of Traumatic Stress, 24*, 609–613. doi:10.1002/jts.20688
- Keane, T. M., Weathers, F. W., & Foa, E. B. (2000). Diagnosis and assessment. In E. B. Foa, T. M. Keane, & M. J. Friedman (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (pp.18–36). New York: Guilford Press.
- Kennedy, C. (2002). SSA's disability determination of mental impairments: A review toward an agenda for research. In G. S. Wunderlich, D. P. Rice, & N. L. Amado (Eds.), *The dynamics of disability: Measuring and monitoring disability for Social Security programs* (pp. 241–280). Washington, DC: National Academies Press.
- Kessler, R. C. (2000). Posttraumatic stress disorder: The burden to the individual and to society. *Journal of Clinical Psychiatry, 61*(Suppl 15), 4–12.
- Miller, P. R. (2001). Inpatient diagnostic assessments: 2. Interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Research, 105*, 265–271. doi:10.1016/S0165-1781(01)00318-3
- Miller, P. R., Dasher, R., Collins, R., Griffiths, P., & Brown, F. (2001). Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Research, 105*, 255–264. doi:10.1016/S0165-1781(01)00317-1
- R Development Core Team. (2006). *R: A language for and environment for statistical computing* [Version 2.10.1]. Vienna, Austria: R Foundation for Statistical Computing.
- Savoca, E., & Rosenheck, R. (2000). The civilian labor market experiences of Vietnam-era veterans: The influence of psychiatric disorders. *Journal of Mental Health Policy and Economics, 3*, 199–207. doi:10.1002/mhp.102
- Schnurr, P. P., Lunney, C. A., Bovin, M. J., & Marx, B. P. (2009). Posttraumatic stress disorder and quality of life: Extension of findings to veterans of the wars in Iraq and Afghanistan. *Clinical Psychology Review, 29*, 727–735. doi:10.1016/j.cpr.2009.08.006
- Shear, M. K., Breeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., & Hanekamp, M. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry, 157*, 581–587. doi:10.1176/appi.ajp.157.4.581
- Sheehan, D. V., LeCrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiler, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.
- Skodol, A. E., Williams, J. B. W., Spitzer, R. L., Gibbon, M., & Kass, F. (1984). Identifying common errors in the use of DSM-III through supervision. *Hospital and Community Psychiatry, 35*(3), 251–255.
- Smith, M. W., Schnurr, P. P., & Rosenheck, R. A. (2005). Employment outcomes and PTSD symptom severity. *Mental Health Services Research, 7*, 89–101. doi:10.1007/s11020-005-3780-2
- Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the Structured Clinical Interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease, 183*, 365–369. doi:10.1097/00005053-199506000-00003
- U.S. Department of Veteran Affairs VA Office of Inspector General. (2006). *Review of state variances in VA disability compensation payments* (05-00765-137 ed.). Washington, DC: Author.
- U.S. Department of Veteran Affairs Veterans Benefits Administration. (2010). *Annual benefits report fiscal year 2010: Making a difference*. Washington, DC: Author.
- Ventura, J., Liberman, R. P., Green, M. F., Shaner, A., & Mintz, J. (1998). Training and quality assurance with the Structured Clinical Interview for DSM-IV (SCID-I/PO). *Psychiatry Research, 70*, 163–173. doi:10.1016/S0165-1781(98)00038-9
- Watson, P. W., McFall, M., McBrine, C., Schnurr, P. P., Friedman, M. J., Keane, T. M., & Hamblen, J. L. (2005). *Best practice manual for posttraumatic stress disorder (PTSD) compensation and pension examinations*. Portland, OR: Northwest Network Mental Illness Research, Education, and Clinical Center, VA Puget Sound Health Care System.
- Weathers, F. W., Keane, T. M., & Davidson, J. R. T. (2001). Clinician-Administered PTSD Scale: A review of the first ten years of research. *Depression and Anxiety, 13*, 132–156. doi:10.1002/da.1029
- Weathers, F. W., Keane, T. M., & Foa, E. B. (2009). Assessment and diagnosis of adults. In E. B. Foa, T. M. Keane, M. J. Friedman, & J. A. Cohen (Eds.), *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies* (2nd ed., pp. 23–61). New York: Guilford Press.
- Weathers, F. W., & Litz, B. T. (1994). Psychometric properties of the Clinician-Administered PTSD Scale, CAPS-1. *PTSD Research Quarterly, 5*, 2–6.
- Weathers, F. W., Newman, E., Blake, D. D., Nagy, L. M., Schnurr, P. P., Kaloupek, D. G., . . . Keane T. M. (2004). *Clinician-Administered PTSD Scale (CAPS)–Interviewer's guide*. Los Angeles, CA: Western Psychological Services.
- World Health Organization. (2004). *Mortality and burden of disease estimates for WHO member states in 2004*. Retrieved from http://www.who.int/entity/healthinfo/global_burden_disease/gbddeathdalycountryestimates2004.xls
- World Health Organization. (2000). *World Health Organization Disability Assessment Schedule WHODAS II. Phase 2 field trials*. Retrieved from <http://www.who.int/entity/healthinfo>
- Zimmerman, M. (2003). What should the standard of care for psychiatric diagnostic evaluations be? *Journal of Nervous Mental Disorders, 191*, 281–286. doi:10.1097/00005053-200305000-00002