Journal of Traumatic Stress
April 2014, *27*, 160–167

ISTSS International Society for Traumatic Stress Studies

Diagnostic Accuracy of the Composite International Diagnostic Interview (CIDI 3.0) PTSD Module Among Female Vietnam-Era Veterans

Rachel Kimerling,[1,2] Tracey Serpi,[3] Frank Weathers,[4] Amy M. Kilbourne,[5,6] Han Kang,[7] Joseph F. Collins,[8] Yasmin Cypel,[7] Susan M. Frayne,[2,9] Joan Furey,[10] Grant D. Huang,[11] Matthew J. Reinhard,[12,13] Avron Spiro,[14,15,16] and Kathryn Magruder[17,18]

[1]National Center for PTSD, VA Palo Alto Health Care System, Menlo Park, California, USA
[2]Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, California, USA
[3]Cooperative Studies Program Coordinating Center, VA Maryland Health Care System, Perry Point, Maryland, USA
[4]Department of Psychology, Auburn University, Auburn, Alabama, USA
[5]VA Ann Arbor Center for Clinical Management Research, Ann Arbor, Michigan, USA
[6]Department of Psychiatry, University of Michigan Medical School, Ann Arbor, Michigan, USA
[7]U.S. Department of Veterans Affairs, Office of Public Health, Post Deployment Strategic Healthcare Group, Washington, DC, USA
[8]Cooperative Studies Program Coordinating Center, VA Maryland Health Care System, Perry Point, Maryland, USA
[9]Division of General Medical Disciplines, Stanford University School of Medicine, Palo Alto, California, USA
[10]Vietnam Veteran, Consultant, Sayville, New York, New York, USA
[11]Cooperative Studies Program Central Office, VA Office of Research & Development, Washington, DC, USA
[12]War Related Illness and Injury Center, Washington, DC VA Medical Center, Washington, DC, USA
[13]Department of Psychiatry, Georgetown University Medical School, Washington, DC, USA
[14]Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, Massachusetts, USA
[15]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA
[16]Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts, USA
[17]Mental Health Service, Ralph H. Johnson VA Medical Center, Charleston, South Carolina, USA
[18]Department of Psychiatry, Medical University of South Carolina, Charleston, South Carolina, USA

The World Health Organization (WHO) Composite International Diagnostic Interview (CIDI) posttraumatic stress disorder (PTSD) module is widely used in epidemiological studies of PTSD, yet relatively few data attest to the instrument's diagnostic utility. The current study evaluated the diagnostic utility of the CIDI 3.0 PTSD module with U. S. women Vietnam-era veterans. The CIDI and the Clinician-Administered PTSD Scale (CAPS) were independently administered to a stratified sample of 160 women, oversampled for current PTSD. Both lifetime PTSD and recent (past year) PTSD were assessed within a 3-week interval. Forty-five percent of the sample met criteria for a CAPS diagnosis of lifetime PTSD, and 21.9% of the sample met criteria for a CAPS diagnosis of past-year PTSD. Using CAPS as the diagnostic criterion, the CIDI correctly classified 78.8% of cases for lifetime PTSD ($\kappa = .56$) and 82.0% of past year PTSD cases ($\kappa = .51$). Estimates of diagnostic performance for the CIDI were sensitivity of .61 and specificity of .91 for lifetime PTSD and sensitivity of .71 and specificity of .85 for past-year PTSD. Results suggest that the CIDI has good utility for identifying PTSD, though it is a somewhat conservative indicator of lifetime PTSD as compared to the CAPS.

This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Clinical Sciences Research and Development, Cooperative Studies Program. The authors would like to thank Allison Rodriguez, Leslie Reinhart, and Craig Kreisler. The authors have no financial involvement with organizations whose financial interest may be affected by the material in the manuscript or which might potentially bias it.

Correspondence concerning this article should be addressed to Rachel Kimerling, VA Palo Alto Health Care System, 795 Willow Road, PTSD-324, Menlo Park, CA 94025. E-mail: Rachel.Kimerling@va.gov

DOI: 10.1002/jts.21905

The World Health Organization (WHO) Composite International Diagnostic Interview (CIDI) posttraumatic stress disorder (PTSD) module is one of the most widely used epidemiological measures of PTSD worldwide. The CIDI version 3.0 (Kessler & Ustun, 2004) was used to identify PTSD and other mental health conditions in the WHO World Mental Health Survey Initiative (Kessler et al., 2007), including the U.S. National Comorbidity Survey and other Collaborative Psychiatric Epidemiological Surveys (Kessler, Chiu, Demler, Merikangas, & Walters, 2005). The CIDI is used internationally for epidemiological studies because it addresses a wide range of mental health conditions, can be administered by trained lay interviewers, and offers a sophisticated computer-assisted telephone interview (or CATI) version. Studies calibrating the CIDI against clinical diagnoses, however, have found only moderate support for its accuracy in identifying PTSD (Haro, 2006; Alegria, 2004; Breslau, Kessler & Peterson, 1998). In addition, no studies to date have assessed the diagnostic performance of the CIDI PTSD module in populations of women or of veterans.

Structured clinical interviews are considered to be a methodologically rigorous diagnostic criterion in evaluation of measures of mental health conditions. Both the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 2002) and the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1995) are structured clinical interviews used widely to generate gold standard diagnoses for PTSD. The concordance between this diagnostic criterion or gold standard diagnosis, and the test diagnosis (i.e., the measure being evaluated) yield estimates of diagnostic accuracy.

Studies of diagnostic accuracy can utilize a variety of statistical indicators. Diagnoses for both the test and criterion measure should be reported to indicate prevalence and to demonstrate that there are sufficient numbers of positive and negative tests to calculate estimates of diagnostic concordance with reasonable precision (Kraemer, 1992). Most studies evaluate concordance using receiver operating curve (ROC) characteristics (Kraemer, 1992; McNeil & Hanley, 1984) such as sensitivity, or "hit rate," the probability that a positive test detects a true diagnosis; and specificity, the probability of a true negative diagnosis. A test with good sensitivity has few false negative diagnoses, where as a test with good specificity has few false positive diagnoses. The diagnostic likelihood ratio (LR) is a function of sensitivity and specificity, but is useful due the ease of interpretation (Sackett, 2009). The LR expresses the certainty of the test diagnosis relative to the diagnostic criterion. An LR ranges from 0 to infinite where values > 1 yield increased certainty of the diagnosis, and values < 1 lend certainty that the diagnosis can be ruled out. The LR+ expresses the likelihood of a positive diagnosis given a positive test, whereas the LR− indicates the likelihood of a negative diagnosis for a negative test result. An LR+ > 5 or LR− < 0.2 suggests strong probability of the condition, where LR+ > 10 or LR− < 0.1 are considered definitive. LRs can then be combined with prevalence to determine the probability of a diagnosis for a positive test (positive predictive value) and the probability the condition is absent with a negative test (negative predictive value). Indicators of overall agreement between a measure and the diagnostic criterion include the area under the ROC curve, an index of the test's ability to discriminate between individuals with and without the disorder, and κ coefficients, the chance corrected probability that the test and the criterion agree. Kappa coefficients have been conceptualized as good indicators of the quality of a test because it is a calibrated indicator of agreement (Kraemer, 1992).

Two studies have examined the diagnostic performance of the CIDI relative to the SCID. In the clinical calibration of a variety of CIDI modules for the World Mental Health Surveys, clinical reappraisal interviews using the SCID were administered to a probability subsample of 325 survey respondents who had participated in the National Comorbidity Survey Reappraisal (NCS-R) within the past 2 months (Haro et al., 2006). The study used a partially unblinded design, whereby respondents were cued to their prior responses to CIDI stem questions. The number of CIDI lifetime PTSD diagnoses was not reported. These data suggested moderate concordance between the CIDI and SCID for lifetime PTSD with a sensitivity of .38 ($SE = .12$), and a specificity of .99 ($SE = .50$), κ of .49 ($SE = .10$) and the area under the ROC curve of .69. The positive predictive value was 86.1% ($SE = 7.7$) and the negative predictive value was 91.3% ($SE = 3.0$). The concordance between past-year CIDI and SCID diagnoses was examined in a sample of 195 U.S. Latinos from the National Latino and Asian American Study (Alegria et al., 2004) using blinded, independent administration of the SCID. Only five cases of PTSD were diagnosed using the SCID, which meant that there were too few criterion diagnoses among the participants to obtain stable estimates of diagnostic accuracy. Of the five diagnosed with PTSD by the SCID, one met criteria for subthreshold PTSD on the CIDI, but none received a PTSD diagnosis with the CIDI. Seven other participants received PTSD diagnoses on the CIDI.

The CAPS is commonly viewed as the standard clinical interview for DSM-IV (4th ed., DSM-IV, American Psychiatric Association, 2000) PTSD (Weiss, 2004). The CAPS is ideal as a diagnostic criterion for PTSD for several reasons: It provides a detailed assessment of both trauma exposure and the frequency and intensity of PTSD symptoms, and it can be scored to obtain a dichotomous indicator of a PTSD diagnosis as well as a continuous measure of symptom severity. To date, we are aware of only one study that compared the CIDI to the CAPS, examining concordance for lifetime PTSD diagnoses. This study was a clinical reappraisal of participants from the Detroit Area Study of Trauma (Breslau, Kessler, & Peterson, 1998), which used a modified version of the CIDI 2.1. In this study, 53 individuals who completed the CIDI 2.1 as part of the original survey were stratified by their survey PTSD diagnosis (32 with lifetime PTSD on the CIDI, 23 without lifetime PTSD on the CIDI, matched on type of trauma exposure) and completed blinded CIDI and CAPS interviews approximately 12–18 months following survey participation. Survey CIDI diagnoses were compared to CAPS diagnoses, yielding a sensitivity of .95 (95% confidence interval [CI] = [.75, 1.0]), specificity of .71, 95%

CI = [.52, .85], and 81% agreement. Weighted estimates of the positive predictive value and the negative predictive value were 75 (95% CI = [30, 97]) and 97 (95% CI = [86, 100]), respectively. Although these data lend greater confidence to the diagnostic accuracy of the CIDI's PTSD module as compared to other studies, additional data on the CIDI 3.0 are needed.

To date, there are limited data that can attest to the diagnostic utility of the CIDI's PTSD module. The prevalence of recent PTSD in the general population is sufficiently low to present a challenge to obtaining sample sizes large enough to estimate diagnostic concordance with a clinical gold standard, and accordingly the diagnostic utility of the CIDI for recent PTSD is unknown (Haro, 2006). The low sensitivity obtained from the NCS-R calibration study despite a partially unblinded design suggests a need for further evaluation. Furthermore, the performance of a diagnostic test may vary across populations, so studies that examine the utility of the CIDI with populations at high risk for PTSD, such as women or veterans may be especially informative. The goal of the current study was to evaluate the diagnostic concordance of the CIDI PTSD module with clinical assessment using the CAPS (Blake et al., 1995) in a sample of women Vietnam-era veterans. We examined the diagnostic certainty of the CIDI PTSD module relative to the CAPS, which has been extensively used to establish clinical diagnoses of PTSD among both veteran and civilian populations (Weathers, Keane, & Davidson, 2001). Diagnostic concordance between the CIDI and CAPS was evaluated for both lifetime and past-year PTSD.

Method

Participants

The study population was drawn from participants who completed a larger epidemiological study of the health status of women veterans who served during the Vietnam-era (Thomas, Kang, & Dalager, 1991) and agreed to participate in the HealthViEWS study. Eligibility criteria for HealthViEWS included the following: women who served as active duty military personnel in one of the four U.S. Armed Services (Army, Navy, Air Force, Marines) between July 4, 1965 and March 28, 1973 with a 30-day minimum period of service in Vietnam, countries near Vietnam, or in the United States. HealthViEWS consisted of a mailed survey followed by a telephone interview, both administered by a survey research company contracted by the Department of Veterans Affairs (VA). The study was approved by the VA Central Institutional Review Board.

Eligibility for the current substudy included return of the mailed survey with complete data for the PTSD Checklist Civilian Version (PCL-C; Weathers, Litz, Herman, Huska, & Keane, 1993), completion of the PTSD module of the CIDI as part of the telephone interview, and agreement to complete an additional telephone interview assessment of PTSD using the CAPS for the purposes of this substudy. All eligible women (1,389) who were approached to participate in the substudy at the conclusion of the telephone interview agreed to participate. One hundred sixty-five women were recontacted and invited to participate in a follow-up interview; the final sample consisted of 160 women who completed the CAPS interview; five participants were excluded due to incomplete CAPS interviews. The average age of women in the sample was 66.7 years ($SD = 4.8$). Demographic characteristics of the sample are described in Table 1.

Table 1
Demographic Characteristics of Participants

Variable	*n*	%
Marital status		
Married or partnered	81	50.6
Divorced or separated	41	25.6
Widowed	5	3.1
Never married	29	18.1
Race		
White	138	86.3
African American	9	5.6
Asian/other	7	4.3
Education		
High school diploma	10	6.3
Some college, vocational school, Associates degree	22	13.8
Three-year degree/diploma	31	19.4
Four-year college degree	40	25.0
Some graduate school	10	6.3
Graduate/professional degree	44	27.5
Employed	127	79.4
Household income		
< $15,000	10	6.3
$15,000–$29,999	15	9.4
$30,000–$49,999	36	22.5
$50,000–$74,999	29	18.1
$75,000–$99,999	24	15.0
$100,000–$149,999	29	18.1
$150,000 or more	8	5.0
PTSD Checklist score		
≥30	68	42.5
<30	92	57.5

Note. $N = 160$. Missing data: Age, marital status, 4; race, 6; education, 3; employment, 3; income, 9. PTSD = posttraumatic stress disorder.

Procedure

The diagnostic concordance between the CIDI and the CAPS was established as part of the Department of Veteran Affairs Cooperative Study #579, an epidemiologic investigation of the health status of women who served during the Vietnam era (HealthViEWS). The purpose was to quantify the diagnostic certainty of the CIDI PTSD module diagnosis in this population by assessing diagnostic concordance with the CAPS. Valid assessment of diagnostic concordance requires

independent, blinded administration of the test diagnosis and criterion, so that results of each assessment do not influence the other, artificially increasing diagnostic concordance, but within a sufficiently short interval of time so that diagnostic concordance is not decreased due to fluctuation in symptoms over time. The current study evaluated 160 women for both past-year and lifetime PTSD using independent, blinded administration of the CIDI PTSD module and CAPS within a 3-week interval.

Representative samples are necessary in studies of diagnostic accuracy, but when the prevalence of a condition is low, sequential or simple random sampling requires very large, often infeasible, sample sizes to estimate agreement statistics with acceptable levels of precision. Smaller samples may yield too few positive diagnoses to estimate accuracy with precision (e.g. Alegria et al., 2004). Under such conditions, balanced samples can be achieved using stratified sampling (Clarke et al., 2013; Knottnerus & Muris, 2003). We employed this sampling approach by prescreening participants for current PTSD symptoms using the PCL-C to try to oversample cases with current PTSD. The PCL-C was completed as part of a HealthViEWS survey prior to study recruitment. Stratified designs must caution against only sampling extreme cases of high or low symptoms and excluding of indeterminate or subthreshold cases. Excluding cases where discrimination is most difficult can upwardly bias estimates of diagnostic accuracy, and is known as spectrum bias. To further guard against spectrum bias, a sensitive cut score ($\geq$ 30; McDonald & Calhoun, 2010; Yeager, Magruder, Knapp, Nicholas, & Frueh, 2007) was used to ensure a wide range of PTSD symptoms was represented, including threshold cases. Of the 160 women in the sample, 68 scored above the PCL-C threshold; 92 scored below. Scores ranged from 17 to 76, with a median PCL-C score of 28 (IQR = 19–43), suggesting threshold cases were represented in the sample.

Women who were eligible for the current substudy were so indicated in the CATI interview system in the HealthViEWS telephone survey. At the conclusion of the telephone survey, women were asked if they would be willing to be contacted to complete a second telephone interview for which they would be compensated an additional $75. Contact information for women who agreed to participate was forwarded to study staff at the Charleston VA Medical Center, and these women were contacted sequentially in blocked recruitment periods for study interviews within 3 weeks of completing the HealthViEWS telephone interview. The CIDI PTSD module was completed during the HealthViEWS telephone interview by trained survey research interviewers blind to survey results, including PCL-C scores. The CAPS assessment was completed independently via telephone by the Charleston VAMC study staff. All CAPS assessments were conducted subsequent to CIDI interviews, but blind to CIDI results and PCL-C.

The CIDI module was administered by trained, experienced lay interviewers who conducted the HealthViEWS telephone interview. Interviewers had between 2 and 20 years experience with a large survey research company and prior experience with surveys that included sensitive topics. In addition to standard training on telephone interviewing, study-specific training, and VA research training, interviewers received substantial instruction on CIDI administration. Supervisors attended intensive CIDI training through the WHO-CIDI 3.0 training program in Ann Arbor, Michigan. The training consisted of a compact disc of take-home lessons that were completed before a 3-day in-person training. During the first 2 days, certified WHO-CIDI trainers provided extensive training on conducting the interview, quality control, and standardized interviewing. On the third day, all attendees were required to participate in a certification interview. Supervisors then provided training to survey research interviewers using CIDI manuals and training materials for self-study, group teleconference trainings, and completion of practice interviews to establish competence. Monitoring by a supervisor was conducted on 10% of all CIDI interviews, with subsequent training provided when necessary.

The CAPS assessment was administered by trained medical clinicians: two doctoral-level psychiatric nurses, a research nurse, and a research physician. All interviewers had a minimum of 5 years of experience in clinical mental health. Two interviewers had extensive prior research experience with CAPS administration. The study coordinator (a doctoral-level psychiatric nurse who also performed interviews) provided a 6-hour interactive refresher training class for CAPS interviews to the interviewers. Three of the first interviews for all four clinicians were recorded. The recordings were scored independently by one or two other interviewers to ensure scoring fidelity.

Measures

Diagnostic interviews for PTSD. The PTSD module of the CIDI version 3.0, CAPI version 20, was used for the diagnosis of past-year and lifetime PTSD (Kessler et al., 2004). The CIDI is a structured interview designed to assess psychiatric disorders according to criteria *based on the DSM-IV* (4th ed.; DSMIV; American Psychiatric Association, 1994) and the *ICD-10: International Statistical Classification of Diseases and Related Health Problems* (World Health Organization, 1992). The PTSD module includes 26 questions to identify potentially traumatic events, and a query for additional events not listed. Participants who report any event are asked additional items to confirm PTSD event criteria of fear, horror, or helplessness. Participants who meet *DSM-IV* trauma exposure criteria are then asked about PTSD symptoms in reference to the worst event (as selected by the participant) and a randomly selected event. There are 17 items with a *yes* or *no* response format, with additional questions about symptom onset and recency. The CIDI was scored using the WHO computerized scoring algorithm corresponding to *DSM-IV* PTSD criteria. The CIDI PTSD module yields information regarding past-year and lifetime PTSD diagnostic status.

The CAPS (Blake et al., 1995) was used as the diagnostic criterion for past-year and lifetime PTSD. The CAPS is a widely used reliable and valid structured clinical interview for PTSD (Weathers et al., 2001). Respondents are first asked about the

occurrence of 20 potentially traumatic events. Those who report any events are then asked additional items to confirm PTSD event criteria of fear, horror, or helplessness in reference to up to three events. Those who meet trauma-exposure criteria are then asked detailed questions about the intensity and frequency of the 17 *DSM-IV* PTSD symptoms in reference to these three events, with standardized probes and clarification. Frequency and intensity of symptoms are each rated on a 0–4 scale. The CAPS is scored according to *DSM-IV* criteria. The most common method for establishing that a symptom is present if the item frequency score is >1 and the intensity score ≥ 2 (F1/I2 scoring rule; Weathers et al., 2001), which was used for the current analyses to facilitate interpretation of results. Because this is also one of the more lenient scoring rules, sensitivity analyses were conducted comparing results obtained with this scoring rule to those obtained with a more-conservative scoring method, using the F1/I2 rule but also requiring a total score of 65 or greater (F1/I2/Sev65; Weathers et al., 2001).

The CAPS traditionally yields a current (past month) and lifetime PTSD diagnosis. The time frame for the CAPS current diagnosis was modified to a past-year diagnosis, to achieve correspondence with the CIDI-PTSD, which yields a past-year diagnosis. This modification was made in collaboration with the primary author of the CAPS, to decrease error in diagnostic concordance due to differing time frames for recent diagnosis between the CIDI and the CAPS. A response prompt was added to CAPS instructions stating "in the past year . . . " to assess symptoms consistent with a past-year diagnosis.

Self-report measures. The PCL-C (Weathers et al., 1993) was used as a self-report measure of PTSD symptoms to stratify recruitment and was included as part of the mail survey. The PCL-C is a 17-item measure of *DSM-IV* PTSD symptoms. Respondents are introduced to "a list of problems and complaints that veterans sometimes have in response to stressful experiences" and are then asked to rate how much each symptom has bothered them during the past month using a 5-point scale (1 = *not at all* to 5 = *extremely*). The total PCL-C score is calculated by summing scores for the 17 items and can range from 17 to 85. The Cronbach's α for the current sample was .93. The PCL has good evidence for its use as a screening tool, especially when used with a sensitive cut score followed by further diagnostic assessment (McDonald & Calhoun, 2010) as in the current study. The PCL-C was selected over the PCL-M for the current study to capture all PTSD symptoms, not only those associated with a military-related traumatic event. Demographic characteristics were also self-reported on the mail survey.

Data Analysis

Analyses were completed using SAS Version 9.2. Concordance between the CIDI and CAPS diagnoses was examined using the CAPS diagnosis as the criterion for both past-year and lifetime PTSD. There were no missing data for these measures. The area under the ROC curve (Hanley & McNeil, 1982) was used as an indicator of CIDI PTSD module accuracy. The area under the ROC curve values range from 0 to 1 where an area under the ROC curve above .5 indicates accuracy greater than chance. We also calculated sensitivity, specificity, and diagnostic likelihood ratios with 95% CIs. Kappa was calculated as an indicator of the chance-adjusted agreement between the two tests. The κ statistic can also be conceptualized as the quality of a diagnostic test's overall discriminative ability (Kraemer, 1992). Similarly, estimates of the quality of sensitivity κ (1.0) and the quality of specificity κ (0.0) were calculated as standardized indicators of sensitivity and specificity. The target range for κ values was .4 to .6. Though traditionally considered a moderate level of agreement (Landis & Koch, 1977), when contextualized within subsequent research into the quality of medical diagnoses, this range has been proposed as a realistic target for agreement across dichotomous diagnoses, where a κ of .6 to .8 would be considered an exceptional quality of agreement (Kraemer, Kupfer, Clarke, Narrow, & Regier, 2012).

Sensitivity analyses were calculated to assess the impact of the CAPS scoring rule, comparing results of the lenient (F1/I2) and more conservative (F1/I2/Sev65) scoring rules. The area under the ROC curve for results from each scoring rule was compared and McNemar tests were used to compare results from each scoring rule on the proportions of true positive CIDI diagnoses, and the proportions of false positive CIDI PTSD module diagnoses for lifetime PTSD (Pepe, 2003). Results revealed no statistically significant differences between estimates obtained from each scoring rule, so all analyses are reported using the more commonly used F1/I2 scoring rule for the CAPS.

Results

The diagnostic concordance of the CIDI with the CAPS is described in Table 2. Criteria for lifetime PTSD according to the CAPS were met by 45.0% of the sample, and 33.7% were diagnosed with PTSD on the CIDI. The CIDI correctly classified 78.8% of cases and κ was .56. Estimates of diagnostic accuracy for the CIDI, using the CAPS as diagnostic criterion yielded a positive LR was 7.03 and negative LR was 0.40.

Criteria for past-year PTSD according to the CAPS were met by 21.9% of the sample, and 27.5% of the sample were diagnosed with PTSD on the CIDI. The CIDI correctly classified 82% of cases; κ was .52. Estimates of diagnostic accuracy for the CIDI, using the CAPS as diagnostic criterion yielded a positive LR was 4.70 and negative LR was 0.34.

We examined misclassified cases by analyzing false negatives and false positives separately. False negative cases were the 26 cases diagnosed with lifetime PTSD by the CAPS, but missed by the CIDI. Among these cases, 11 (42%) demonstrated subthreshold levels of PTSD symptoms on the CIDI, meeting criteria for partial PTSD (Stein, Walker, Hazen, & Forde, 1997), defined as fewer than the required number of symptoms for either Criteria C or D, or failed to meet only the E or F criteria.

Table 2
Estimates of Diagnostic Accuracy for the CIDI Compared to the CAPS as Diagnostic Criterion for PTSD

	CAPS		CIDI							
Variable	*n*	%	*n*	%	AUC	95% CI	Sens	95% CI	Spec	95% CI
Lifetime PTSD	72	45%	54	33.7%	0.77	[0.71, 0.84]	0.64	[0.52, 0.75]	0.91	[0.83, 0.96]
Past-year PTSD	35	21.9%	44	27.5%	0.78	[0.70, 0.86]	0.71	[0.54, 0.85]	0.85	[0.77, 0.91]
	LR+	95% CI		LR−		95% CI	κ	95% CI	κ (1)	κ (0)
Lifetime PTSD	7.03	[3.55, 13.92]		0.40		[0.29, 0.54]	0.56	[0.43, 0.69]	0.45	0.73
Past-year PTSD	4.70	[2.96, 7.47]		0.34		[0.20, 0.57]	0.52	[0.36, 0.67]	0.61	0.45

Note. PTSD = posttraumatic stress disorder; CAPS = Clinician Administered PTSD Scale; CIDI = Composite International Diagnostic Interview; AUC = area under the receiver operating characteristic curve; Sens = sensitivity; Spec = specificity; LR+ = positive likelihood ratio; LR− = negative likelihood ratio; κ(1) = quality of the sensitivity; κ(0) = quality of the specificity.

Five cases did not meet trauma-exposure criteria on the CIDI, and were not assessed for PTSD symptoms due to skip patterns.

False positives were the eight cases that were not diagnosed with lifetime PTSD by the CAPS, but met criteria for lifetime PTSD on the CIDI. Among these cases, five demonstrated subthreshold levels of PTSD on the CAPS, meeting criteria for all but one symptom cluster.

Discussion

This study supports the use of the CIDI PTSD module for identification of PTSD among women veterans of the Vietnam era. Using CAPS as the diagnostic criterion, the CIDI correctly classified approximately 80% of cases (79% lifetime PTSD; 82% past-year PTSD), with κ coefficients for diagnostic concordance well within the target range. The clinical significance of the area under the ROC curve, as a measure of diagnostic accuracy (.77 for lifetime, .78 for past-year) indicates a very large effect size (Kraemer et al., 2003), lending further support to the CIDI as a diagnostic tool for PTSD. The strengths of this study, including independent, blinded assessments, both assessments occurring within 3 weeks of each other, a sensitive and detailed reference standard, and a sample size sufficient to examine both lifetime and past-year PTSD, lend further confidence to our data, suggesting that the CIDI 3.0 PTSD module is an acceptable diagnostic tool for PTSD among female Vietnam-era veterans, and a promising instrument for use with other populations.

A clinician-administered diagnostic interview with the sensitivity and detail that the CAPS provides is not feasible for large-scale epidemiological studies, making the CIDI an attractive alternative. Even a comprehensive and structured lay-administered interview, however, such as the CIDI cannot be expected to achieve a level of precision similar to a clinical interview. Some measurement error in the epidemiological diagnosis of PTSD using the CIDI is therefore unavoidable, though in this case the limitation is balanced by the fact that use of the CIDI permits a large sample size and external validity rarely afforded by clinical studies. The CIDI appears to be a somewhat conservative indicator of PTSD, demonstrating better specificity (ability to rule out negative cases) as compared to sensitivity (the ability to detect positive cases), for lifetime diagnoses. This direction of misclassification bias is appropriate: An overly sensitive instrument would bias away from the null hypothesis, and inflate both prevalence estimates and the ability to detect statistically significant associations. A more-specific instrument biases results towards the null hypothesis, and increases confidence in effects that do achieve statistical significance.

Our study found stronger support for the CIDI's diagnosis of PTSD as compared to evaluations of diagnostic accuracy in the WHO World Mental Health Surveys (Alegria et al., 2009; Haro et al., 2006; Kraemer et al., 2003), especially with respect to the instrument's sensitivity. Our accuracy results are more comparable to those using version 2.1 of the CIDI PTSD module (Breslau et al., 1998), possibly because both studies were powered specifically to examine PTSD diagnoses. Breslau and colleagues did find, in contrast to our results, the CIDI to be more sensitive than specific in the diagnosis of lifetime PTSD. In both the Breslau et al. (1998) study and the current study, however, the majority of false positive cases on the CIDI were identified as subthreshold cases of PTSD by the CAPS. These results lend further confidence to PTSD diagnoses established by the CIDI. Correspondingly, the largest group of false negative diagnoses by the CIDI also met criteria for subthreshold PTSD. A substantial literature has addressed the common frequency and clinical significance of partial or subthreshold *DSM-IV* PTSD (Grubaugh et al., 2005; Marshall et al., 2001; Pietrzak, Goldstein, Southwick, & Grant, 2011), and for such cases there will always be some degree of diagnostic uncertainty in the administration of any single measure.

This study should be interpreted in light of several considerations. Most important, though promising, our results do not necessarily generalize to other populations, and additional research is needed to determine the specific psychometric properties of the CIDI 3.0 PTSD module with males, nonveterans, and other age groups. The short interval of time between the CIDI interview and the CAPS assessment is a strength of the study,

but because we did not vary the order of administration of the CIDI and the CAPS, participants may have been primed by the CIDI interview for traumatic events and PTSD symptoms, resulting in enhanced event and symptom reporting for the CAPS. Finally, as noted by Haro et al. (2006), inferences about the validity of any diagnostic interview must be considered in the context of measurement error in the diagnostic gold standard. The CAPS test-retest reliability of .89 (Weathers et al., 2001) is very good, but as with all measures, imperfect. Our results should therefore be interpreted as a lower-bound estimate of CIDI accuracy.

In summary, our results support the diagnostic utility of the CIDI PTSD module among female Vietnam-era veterans and provide incremental support, along with Breslau and colleagues (1998), for the diagnostic utility of the CIDI. Although future research should continue to examine the CIDI in trauma populations of interest, we suggest that the CIDI PTSD module appears to be an acceptable epidemiological tool to detect PTSD among women veterans.

References

Alegria, M., Shrout, P., Torres, M., Lewis-Fernndez, R., Abelson, J., Powell, M., . . . Canino, G. (2009). Lessons learned from the clinical reappraisal study of the Composite International Diagnostic Interview with Latinos. *International Journal of Methods in Psychiatric Research*, *18*, 84–95. doi:10.1002/mpr.280

Alegria, M., Takeuchi, D., Canino, G., Duan, N., Shrout, P., Meng, X. L., . . . Gong, F. (2004). Considering context, place and culture: The National Latino and Asian American Study. *International Journal of Methods in Psychiatric Research*, *13*, 208–220.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a Clinician-Administered PTSD Scale. *Journal of Traumatic Stress*, *8*, 75–90. doi:10.1002/jts.2490080106

Breslau, N., Kessler, R., & Peterson, E. L. (1998). Post-traumatic stress disorder assessment with a structured interview: Reliability and concordance with a standardized clinical interview. *International Journal of Methods in Psychiatric Research*, *7*, 121–127. doi: 10.1002/mpr.41

Clarke, D. E., Narrow, W. E., Regier, D. A., Kuramoto, S. J., Kupfer, D. J., Kuhl, E. A., . . . Kraemer, H. C. (2013). *DSM-5* field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry*, *170*, 43–58. doi:10.1176/appi.ajp.2012.12070998

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV Axis I Research Version, Non-patient Edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.

Grubaugh, A. L., Magruder, K. M., Waldrop, A. E., Elhai, J. D., Knapp, R. G., & Frueh, B. C. (2005). Subthreshold PTSD in primary care: Prevalence, psychiatric disorders, healthcare use, and functional status. *Journal of Nervous and Mental Disease*, *193*, 658–664. doi:10.1097/01.nmd.0000180740.02644.ab

Hanley, J. A., & McNeil, B. J. (1982). Maximum attainable discrimination and the utilization of radiologic examinations. *Journal of Chronic Diseases*, *35*, 601–611.

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., De Girolamo, G., Guyer, M. E., Jin, R., . . . Kessler, R. C. (2006). Concordance of the Composite International Diagnositc Interview version 3.0 (CIDI 3.0) with standardised clinical assessments in the WHO World Mental Health Surveys. *International Journal of Methods in Psychiatric Research*, *14*, 167–180. doi:10.1002/mpr.196

Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., . . . Zheng, H. (2004). Clinical calibration of *DSM-IV* diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 122–139. doi:10.1002/mpr.169

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasquet, I., . . . Ustun, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*, 168–176.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month *DSM-IV* disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617–627. doi:10.1001/archpsyc.62.6.617

Kessler, R. C., & Ustun, T. B. (2004). The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 93–121.

Knottnerus, J. A., & Muris, J. W. (2003). Assessment of the accuracy of diagnostic tests: The cross-sectional study. *Journal of Clinical Epidemiology*, *56*, 1118–1128. doi:10.1016/S0895-4356(03)00206-3

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

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). *DSM-5*: How reliable is reliable enough? *American Journal of Psychiatry*, *169*, 13–15. doi:10.1176/appi.ajp.2011.11010050

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 1524–1529. doi:10.1097/01.chi.0000091507.46853.dl

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Marshall, R. D., Olfson, M., Hellman, F., Blanco, C., Guardino, M., & Struening, E. L. (2001). Comorbidity, impairment, and suicidality in subthreshold PTSD. *The American Journal of Psychiatry*, *158*, 1467–1473. doi:10.1176/appi.ajp.158.9.1467

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review*, *30*, 976–987. doi:10.1016/j.cpr.2010.06.012

McNeil, B. J., & Hanley, J. A. (1984). Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Medical Decision Making*, *4*, 137–150. doi:10.1177/0272989X8400400203

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction* (Vol. *31*). New York, NY: Oxford University Press.

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Medical comorbidity of full and partial posttraumatic stress disorder in US adults: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychosomatic Medicine*, *73*, 697–707. doi:10.1097/PSY.0b013e3182303775

Sackett, D. L. (2009). A primer on the accuracy of the clinical examination. In D. Simel & D. Remmie (Eds.), *The rational clinical examination*. New York, NY: McGraw-Hill Education.

Stein, M. B., Walker, J. R., Hazen, A. L., & Forde, D. R. (1997). Full and partial posttraumatic stress disorder: Findings from a community survey. *American Journal of Psychiatry*, *154*, 1114–1119.

Thomas, T. L., Kang, H. K., & Dalager, N. A. (1991). Mortality among women Vietnam veterans, 1973–1987. *American Journal of Epidemiology*, *134*, 973–980.

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-administered PTSD scale: A review of the first ten years of research. *Depression & Anxiety*, *13*, 132–156. doi:10.1002/da.1029

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the Annual Meeting of International Society for Traumatic Stress Studies, San Antonio, TX.

Weiss, D. (2004). Structured Clinical Interview Techniques for PTSD. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 103–121). New York, NY: Guilford Press.

World Health Organization. (1992). *ICD-10: International statistical classification of diseases and related health problems*. Geneva, Switzerland: Author.

Yeager, D. E., Magruder, K. M., Knapp, R. G., Nicholas, J. S., & Frueh, B. C. (2007). Performance characteristics of the posttraumatic stress disorder checklist and SPAN in Veterans Affairs primary care settings. *General Hospital Psychiatry*, *29*, 294–301. doi:10.1016/j.genhosppsych.2007.03.004