

Validation of lay-administered mental health assessments in a large Army National Guard cohort

MARTA R. PRESCOTT,^{1,4} MARIJO TAMBURRINO,² JOSEPH R. CALABRESE,³ ISRAEL LIBERZON,⁴
RENEE SLEMBARSKI,³ EDWIN SHIRLEY,³ THOMAS FINE,² TOYOMI GOTO,³ KIMBERLY WILSON,²
STEPHEN GANOCY,³ PHILIP CHAN,³ ALPHONSE DERUS,³ MARY BETH SERRANO,³
JAMES SIZEMORE,⁵ JEREMY KAUFFMAN⁵ & SANDRO GALEA^{1,4}

1 Columbia University, New York, USA

2 University of Toledo College of Medicine, Toledo, OH, USA

3 Department of Psychiatry, University Hospitals Case Western Reserve University, Cleveland, OH, USA

4 University of Michigan, Ann Arbor, MI, USA

5 Columbus Veterans Medical Center, Columbus, OH, USA

Key words

military, assessment, “posttraumatic stress disorder”, “depressive disorders”, “alcohol use disorders”

Correspondence

Sandro Galea, Mailman School of Public Health, Columbia University, 722 168th St. 1508, New York, USA.
Telephone (+1) 212-305-8755
Email: sg822@columbia.edu

Received 11 April 2012;
revised 7 December 2012;
accepted 16 January 2013

Abstract

To report the reliability and validity of key mental health assessments in an ongoing study of the Ohio Army National Guard (OHARNG). The 2616 OHARNG soldiers received hour-long structured telephone surveys including the post-traumatic stress disorder (PTSD) checklist (PCV-C) and Patient Health Questionnaire – 9 (PHQ-9). A subset ($N=500$) participated in two hour clinical reappraisals, using the Clinician-Administered PTSD Scale (CAPS) and the Structured Clinical Interview for DSM (SCID). The telephone survey assessment for PTSD and for any depressive disorder were both highly specific [92% (standard error, SE 0.01), 83% (SE 0.02)] with moderate sensitivity [54% (SE 0.09), 51% (SE 0.05)]. Other psychopathologies assessed included alcohol abuse [sensitivity 40%, (SE 0.04) and specificity 80% (SE 0.02)] and alcohol dependence [sensitivity, 60% (SE 0.05) and specificity 81% (SE 0.02)]. The baseline prevalence estimates from the telephone study suggest alcohol abuse and dependence may be higher in this sample than the general population. Validity and reliability statistics suggest specific, but moderately sensitive instruments. Copyright © 2014 John Wiley & Sons, Ltd.

Introduction

The link between combat exposure and psychopathologies, including post-traumatic stress disorder (PTSD), depression, anxiety, and substance abuse among military populations is well documented (Killgore *et al.*, 2006; Johnson *et al.*, 2009).

Studies suggest that between 4.8–18% of military populations have had PTSD at some point in their lifetimes (Hoge *et al.*, 2004; Dohrenwend *et al.*, 2006; Vasterling *et al.*, 2006; Iversen *et al.*, 2009) compared with a 6.8–9.2% lifetime prevalence of PTSD for the general US population (Breslau *et al.*, 1998; Kessler *et al.*, 2005a). Similarly, studies suggest that military

personnel have a greater lifetime prevalence of depression and generalized anxiety compared with the general population (Hoge *et al.*, 2004; Kulka, 1990).

During Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF) the National Guard and Reserve forces were deployed to combat zones at an unprecedented level (Vogt *et al.*, 2008). Little is understood about the long-term effects of deployment on National Guard soldiers compared to their active duty counterparts. Some studies suggest that Guard soldiers may be at greater risk of deployment stressors and adverse mental health effects of war than active duty soldiers (Smith *et al.*, 2008). For example, Guard soldiers deployed to conflict areas are exposed to the same combat experiences as active duty personnel but face different deployment stressors including maintaining a civilian job while deployed and deploying with a unit with which they did not train (Hotopf *et al.*, 2006; Vogt *et al.*, 2008; La Bash *et al.*, 2009). Additionally, National Guard veterans face different stressors upon returning home including limited access to health care compared to active duty soldiers (Milliken *et al.*, 2007). Milliken *et al.* (2007) screened soldiers six months after their return from Iraq and found that, compared with active duty forces, twice as many reserve members required referral for mental health problems. As we approach the end of OEF and OIF and given the lack of understanding about how deployment affects reserve forces over time, there is a need to document mental health over time in the National Guard population.

Assessment of mental health conditions by trained clinicians is considered the gold standard but is costly and logistically challenging within large population-based studies (Smith *et al.*, 2007a). As a result, cohort studies of mental health have historically employed more practical interview methods including web-based self-report surveys as conducted by the Millennium Cohort, a large US military cohort (Smith *et al.*, 2007a, 2007b), or telephone-based interviews as conducted by the Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (Remington *et al.*, 1988).

The Ohio Army National Guard Mental Health Initiative (OHARNG MHI) is a longitudinal study that annually monitors the factors associated with and course of mental health within a representative sample of service members from the OHARNG (Calabrese *et al.*, 2011; Goldmann *et al.*, 2012). We report here the psychometrics of the structured mental health assessments completed with a computer-assisted telephone interview (CATI) as compared to the gold standard of clinical face-to-face interviews.

Methods

Study population and sampling

The study population of the OHARNG MHI is the OHARNG soldiers who served in the Guard between June 2008 and February 2009; the final study sample is 2616 randomly selected OHARNG soldiers [men and women, 18 years or older (with some 17 year old emancipated minors) of any ethnicity and capable of informed consent]. OHARNG soldiers were invited to participate through a process that included, first, a letter alerting soldiers of the study with an option to opt-out and, second, a telephone call to obtain each soldier's consent to participate in a telephone interview.

During the first stage of enrollment, all soldiers enlisted in the OHARNG between June 2008 and February 2009 received alert-letters directly from the OHARNG ($N = 12,225$ excludes the 345 without an address). Of all guard soldiers who received the alert-letters, 8.2% (1013 soldiers) returned opt-out cards to the OHARNG.

During the second stage of enrollment, we contacted possible participants to obtain informed consents for the telephone interviews. If the service member was deployed at the time of contact, information was requested on when the member would return and a call was scheduled. If after 10 telephone calls for a two week period at different times of the day and contact was unsuccessful, a non-contact letter was sent to the possible participant's address in an attempt to obtain a working telephone number.

The consent procedure and survey were piloted in November 2008 with 15 service members using a CATI. Official enrolment began in December 2008 and continued through the end of November 2009 when the desired sample size was reached. Participants were compensated for their time.

Clinical reappraisal

We also conducted clinical reappraisals on a sub-sample of the telephone survey participants. At the end of the initial telephone interviews, a random sample of 500 participants participated in the in-depth clinical interview. In-person interviews, conducted by Doctoral and Masters level clinicians, took place in a setting familiar to the participant, averaged two hours, and participants were compensated for their time.

Assessment instruments

The OHARNG MHI CATI included questions on lifetime experiences, deployment and military experiences, current living situation, and past and present symptoms of psychopathology.

Psychopathologies were assessed using standardized and well-validated scales. The PTSD Checklist (PCL-C) (Blanchard *et al.*, 1996) was used to collect PTSD symptoms in relation to participants' self-identified "worst" event experienced both outside and during their most recent deployments (Blake *et al.*, 1995; Weathers *et al.*, 1999; Hoge *et al.*, 2004). Questions were added to assess additional criteria for PTSD diagnosis as listed in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association, 2000). To have had PTSD, a person had to experience criterion A1 and A2 (experience a traumatic event and intense fear, hopelessness or horror due to a trauma); criterion B where at least one symptom of re-experiencing the trauma was reported; criterion C where at least three symptoms of avoidance of the trauma were reported; criterion D where at least two symptoms of increased arousal were reported; criterion E where symptoms lasted for at least one month; and criterion F where the symptoms caused significant impairment (Weathers *et al.*, 1999; American Psychiatric Association, 2000). For non-deployment, deployment, or total PTSD, all symptoms had to be related to one self-identified "worst" trauma either outside the most-recent deployment, during the most-recent deployment, or at any time, respectively.

To assess depressive episodes and obtain occurrence of suicidal ideation, we used the Primary Care Evaluation of Mental Health Disorders Patient Health Questionnaire – 9 (PHQ-9) (Kroenke *et al.*, 2001). To have had a major depressive disorder (MDD), the participant had to report ≥ 5 of nine symptoms on the PHQ-9 and symptoms had to occur together within a two-week period along with either depressed mood or anhedonia. We also examined a more inclusive definition of depression defined by those who, within a two week period with either depressed mood or anhedonia, scored ≥ 2 out of nine symptoms on the PHQ-9 (Kroenke *et al.*, 2001). Suicidal ideation was assessed through the PHQ-9 question asking whether participants had thoughts of death or wanting to hurt themselves within the past 30 days (Kroenke *et al.*, 2001).

Generalized anxiety disorder (GAD) was assessed with the Generalized Anxiety Disorder – 7 (GAD-7) (Spitzer *et al.*, 2006). A probable case of GAD was classified as a score ≥ 10 on the GAD-7, duration of symptoms at least six months, reported functional impairment, with symptoms grouped together (Spitzer *et al.*, 2006). As the clinical reappraisal interview only captured current cases, we only examined current cases of GAD in the past 30 days.

The Mini International Neuropsychiatric Interview (MINI) and DSM-IV criteria were used to assess alcohol dependence and alcohol abuse (Sheehan *et al.*, 1998). Participants with lifetime alcohol abuse ever in lifetime

met DSM-IV criterion 1 (at least one symptom of maladaptive pattern of substance use leading to clinically significant impairment or distress) and criterion 2 (symptoms never met the criteria for alcohol dependence) (Sheehan *et al.*, 1998; American Psychiatric Association, 2000). Those with alcohol dependence ever in lifetime met at least three symptoms of maladaptive pattern of alcohol use leading to clinically significant impairment or distress (Sheehan *et al.*, 1998; American Psychiatric Association, 2000).

Clinical reappraisal instruments

For the clinical reappraisal, the Clinician-administered PTSD Scale (CAPS) was used to assess PTSD based on the "worst" event outside of their deployments as well as the "worst" event during any deployment; deployment events were not limited to the most recent deployment as with the telephone interview (Blake *et al.*, 1995; Weathers *et al.*, 1999). The diagnosis of PTSD for the clinical reappraisal was based on the scoring rules outlined by Weathers *et al.* (1999) for the CAPS and followed the DSM-IV algorithm (Blake *et al.*, 1995; Weathers *et al.*, 1999; American Psychiatric Association, 2000). To have a positive symptom for DSM-IV PTSD criteria B–D, a participant had to have a frequency ≥ 1 per symptom (at least once or twice in their lifetime) as well as a symptom intensity of ≥ 2 (at least moderate – distress clearly present but still manageable and some disruption of activities). To be diagnosed with PTSD a participant had to have all criteria from the DSM-IV (criteria A–F).

The diagnoses for lifetime occurrence of MDD, alcohol abuse and alcohol dependence, and current occurrence of GAD were based on the Structured Clinical Interview for DSM-IV-TR (SCID) Axis I Disorders (non-patient version) and DSM-IV criteria (American Psychiatric Association, 2000).

Suicidal ideation was evaluated using MINI Plus (Sheehan *et al.*, 1998). A positive response was a score of at least "moderately" (nine points or greater) on the question of suicide attempts in the past six months.

Statistical methods

First, we compared the distribution of demographic characteristics (e.g. age, gender, and education) from those in the baseline sample [telephone survey ($N = 2616$)] and those later selected to participate in the clinical reappraisal ($N = 500$) using chi-square tests.

Second, the lifetime prevalence of each psychopathology – PTSD, MDD, any depressive disorder, GAD (past 30 days), alcohol abuse, alcohol dependence, and suicidal ideation (past 30 days) – was described for the entire telephone survey sample.

Third, we examined the validity and reliability of the telephone assessments compared with the clinical reappraisal. Using the 500 participants who were in both samples, we applied four tests of validity and three tests of reliability following methods presented by Kessler *et al.* (2005b) in the National Co-morbidity Survey Replication (NCS-R) (Kessler *et al.*, 2005b).

To assess validity and using the clinical reappraisal as the gold standard, we calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for all psychopathologies. Next, using the overall continuous score from each of the psychopathology scales, we examined the area under the curve (AUC) as a measure of overall accuracy based on the continuous score of the telephone assessment and the gold standard of the clinical interview. All standard errors reported were asymptotic.

To assess reliability, we calculated the kappa statistic and the McNemar's statistic between diagnoses according to the telephone interview and clinical reappraisal. The final measure of reliability was Cronbach's alpha applied to the telephone survey questions.

Finally, to test whether disease misclassification between the telephone and the clinical reappraisal depended on participant characteristics, we compared the sensitivity and specificity for each psychopathology calculated separately for men and women, participants < 35 and ≥ 35 years of age, and White and non-White categories. Confidence intervals (CIs) for these statistics were asymptotic unless the sample size was ≤ 50, in which case exact CIs were reported.

All analyses were carried out with SAS 9.2.

Results

Of the 11,212 soldiers for whom contact information was received from the Guard, 10.1% (1130) were excluded because they did not have a listed telephone number or address and 31.8% (3568) were excluded due to non-functioning or incorrect numbers and not returning a non-contact letter (Figure 1). Of the 6514 possible participants with working numbers (58.1% of the original telephone number list), only 20.9% (1364) declined to participate and 36.0% (2347) were not included because they were not enrolled before the baseline cohort closed in November 2009 ($n=2316$) or disqualified for other reasons (e.g. did not speak English, hearing problems, or deceased, $n=31$), 187 were retired and therefore ineligible. Overall, our participation rate was 43.2% calculated as those who completed the telephone survey plus those who would have consented had they not been retired divided by all of the working numbers minus those disqualified for other reasons.

There were no differences between the characteristics of the baseline and clinical reappraisal samples (Table 1). The majority of participants were male (85.2%), White (87.8%),

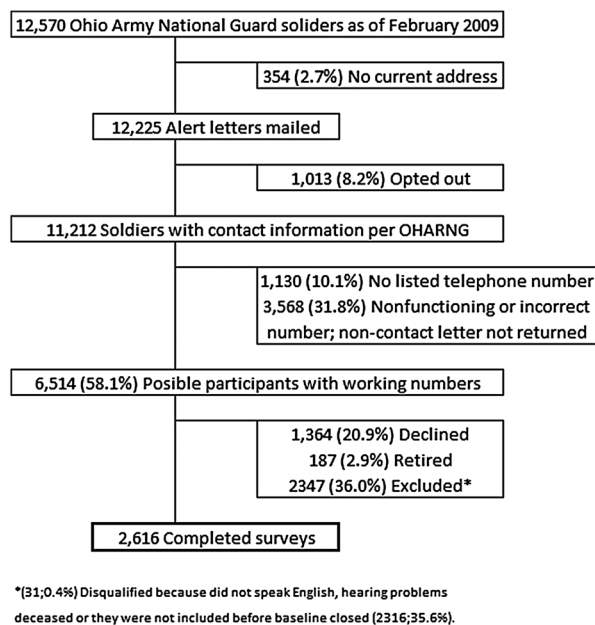


Figure 1. The 2616 completed surveys.

and non-officers, including enlisted soldiers, cadets, or civilian employees (86.9%). The majority had some form of deployment/mobilization experience (36.1% never deployed in the baseline sample); 30.5% of the sample were most recently deployed to a conflict setting.

Table 2 lists the prevalence of each condition in the total baseline sample. The most commonly reported lifetime condition was alcohol abuse (24.0%) followed by alcohol dependence (23.5%). Of the sample, 10.3% had MDD at some point in their lives and 21.4% had some form of depression (MDD including other forms of depression). Deployment-related PTSD was reported by 9.6% of the telephone sample while 10.1% had PTSD ever in lifetime. GAD (1.7%) and suicide risk (1.9%) were rarely reported.

For the validity measures (Table 3), specificity and NPV were higher than sensitivity and PPV for all diagnoses. The telephone diagnosis was most sensitive for alcohol dependence (0.60) and least sensitive for GAD (0.04). The telephone diagnosis was most specific for GAD (0.98) and least specific for alcohol abuse (0.80). The PPV varied but was moderate to low for all conditions, the highest being for MDD (0.64). The NPV was very high for all conditions, the lowest being for alcohol abuse (0.77). Reliability statistical testing results (Table 3) produced relatively moderate kappa values, for example, 0.34 for PTSD ever in lifetime and 0.37 for alcohol dependence. McNemar's test rejected the null hypothesis of no marginal heterogeneity between the telephone sample and clinical interview sub-sample for PTSD, MDD, GAD, and alcohol dependence. The measure of

Table 1. Participant characteristics: telephone interview sample compared with clinical interview sub-sample

Variable ¹	Telephone interview		Clinical interview		P-Value ²
	N=2616		N=500		
	<i>n</i>	%	<i>n</i>	%	
Sex					
Men	2228	85.2	440	88.0	0.10
Women	388	14.8	60	12.0	
Age, years					
17–24 ³	878	33.6	160	32.0	0.14
25–34	848	32.5	182	36.4	
35–44	634	24.3	103	20.6	
≥45	250	9.6	55	11.0	
Race					
White	2295	87.8	444	88.8	0.73
Black	195	7.5	35	7.0	
Other	123	4.7	20	4.0	
Income					
≤ \$60,000	1498	59.1	279	55.8	
> \$60,001	1038	40.9	205	38.0	
Education					
High school graduate/GED or less	727	27.8	137	27.4	0.94
Some college or technical training	1234	47.2	240	48.0	
College/graduate degree	655	25.0	123	24.6	
Marital status					
Married	1227	47.0	238	47.6	0.71
Divorced/separated/widowed	252	9.6	53	10.6	
Never married	1134	43.4	209	41.8	
Rank					
Officer	342	13.1	56	11.2	0.25
Enlisted/cadet/civilian employee	2273	86.9	444	88.8	
Most recent deployment location					
Never deployed	939	36.1	173	34.6	0.66
Non-conflict area	872	33.5	178	35.6	
Conflict area	793	30.5	146	29.2	
Number of lifetime deployments					
0–1	1756	67.4	323	64.6	0.49
2–3	682	26.2	143	28.6	
≥ 4	169	6.5	30	6.0	
Total number of traumatic events experienced					
0	141	5.4	23	4.6	0.65
1–5	887	33.9	159	31.8	
6–11	831	31.8	166	33.2	
≥12	757	28.9	152	30.4	

¹Some percentages do not equal 100% because of missing values.

²Chi-square tests.

³Emancipated minors as defined by Ohio state law were eligible.

reliability and internal agreement for the telephone psychopathologies reported by Cronbach's alpha ranged from 0.95 for deployment-related PTSD to 0.57 for alcohol abuse.

The sensitivity and specificity of the telephone diagnoses stratified by gender, age, and race across the psychopathologies showed no misclassification related to these demographic

Table 2. Prevalence of mental health conditions for telephone interview sample

Disorder	Telephone interview	
	Total (N=2616)	
	N	%
Alcohol abuse ¹	628	24.0
Alcohol dependence ²	615	23.5
Major depressive disorder ³	270	10.3
Any depressive disorder ⁴	560	21.4
Deployment-related PTSD ^{5,6}	121	9.6
PTSD ever in lifetime ^{6,7}	249	10.1
Generalized anxiety disorder ⁸	45	1.7
Suicide risk ⁹	49	1.9

¹DSM-IV criterion A (at least one symptom of maladaptive pattern of substance use leading to impairment or distress) and criterion B (does not meet requirements for substance dependence ever in lifetime). Those who reported never having drunk were coded as never having the condition.

²DSM-IV criterion A (at least three symptoms of maladaptive pattern of substance use) ever in lifetime and symptoms occurred together; MINI. Those who reported never having drunk were coded as never having the condition.

³DSM-IV criteria; ≥ 5 out of nine on PHQ-9, depressed mood or anhedonia, and symptoms occurred together.

⁴DSM-IV criteria; ≥ 2 out of nine on PHQ-9, depressed mood or anhedonia, and symptoms occurred together.

⁵Calculated among everyone who have deployment experience (N=1668) minus those who never experienced a deployment related traumatic event (N=374) and those who refused to answer deployment-related PTSD symptoms (N=28). Of the total sample, nine individuals refused to say if they had ever been deployed and were coded as missing.

⁶DSM-IV criterion A/A2 and criteria B-F ever in lifetime.

⁷Calculated among everyone in the sample (N=2616) minus those who never experienced a traumatic event (N=141) and those who refused to answer PTSD symptoms (N=14).

⁸ ≥ 10 on GAD-7, at least six months symptom duration, functional impairment, symptoms occurred together, and presence of symptoms in the past month.

⁹PHQ-9 (thoughts of wanting to hurt themselves in the past 30 days).

variables (Tables 4 and 5). There was evidence of misclassification for alcohol abuse by gender; the sensitivity and specificity for alcohol abuse was higher for men than women.

Discussion

Overall, the validity and reliability statistics for the computer-assisted telephone psychopathology assessment indicated that

the methods performed well as research instruments for research on PTSD, depression, alcohol abuse, and suicide risk.

All structured screening instruments had high specificity, a necessary characteristic in order to accurately estimate population prevalences (Terhakopian *et al.*, 2008). The sensitivity and specificity for nearly all of the psychopathology diagnoses in the telephone sample did not differ by demographic group, suggesting there was no differential misclassification. This implies that any misdiagnoses for these conditions are random, rather than based on participant characteristics. There was, however, some suggestion that alcohol abuse may be misclassified by gender; women were less likely to be correctly diagnosed than men. Given the high specificity and moderate sensitivity, telephone assessments will be particularly important in the long-term for population assessments and research tools. They will not, however, replace traditional methods of screening for individual treatment.

The telephone assessments had moderate to high levels of reliability across the three measures assessed: kappa, Cronbach's alpha, and McNemar's test. The kappa statistics were fair for suicide risk and all diagnoses with the exception of GAD, suggesting that agreement between the telephone and clinical diagnoses was not due to chance, other than possibly for GAD (Table 3). However, the statistics for GAD showed good internal consistency. Spitzer (2006) reported a Cronbach alpha of 0.92 for his GAD validation study, higher than ours (0.72), but comparable (Spitzer *et al.*, 2006). The other Cronbach alphas in Table 3 also indicate consistency and that the index questions represented the same underlying construct.

Lastly, for McNemar's test of reliability, the finding that psychopathology diagnostic results for several conditions did not reject the null of marginal homogeneity suggested that the telephone assessment and clinical interview were using the same core criteria for diagnoses of alcohol abuse, any depressive disorder, and suicide risk. In comparison, PTSD, MDD, GAD, and alcohol dependence tests rejected the null of marginal homogeneity, suggesting some differences in the core diagnostic criteria between the telephone and the clinical interview sub-sample. As the MDD diagnosed on the telephone compared to the clinical interview varied, we compared the general depression (including MDD and other forms of depression) prevalence from the telephone sample with MDD in the clinical interview sub-sample. We found these two diagnostic tests were more reliable and appeared to use the same diagnostic criteria. It is of note that in the NCS-R, Kessler *et al.* (2005b) reported comparable reliability statistics for these psychopathologies. However, Kessler *et al.* (2005b) found core diagnostic differences by McNemar's test between the

Table 3. Reliability and validity statistics for the telephone assessment sample by psychopathology

Disorder ¹	Sensitivity (SE) ²	Specificity (SE) ³	PPV (SE) ⁴	NPV (SE) ⁵	Kappa (SE) ⁶	McNemar ⁷	Cronbach's alpha (standardized) ⁸	AUC ⁹
Deployment-related PTSD	0.50 (0.13)	0.93 (0.02)	0.35 (0.11)	0.97 (0.01)	0.36 (0.11)	1.8	0.95	0.71
Non-deployment-related PTSD	0.47 (0.13)	0.94 (0.01)	0.23 (0.08)	0.98 (0.01)	0.27 (0.09)	8.0*	0.93	0.74
Any PTSD	0.54 (0.09)	0.92 (0.01)	0.31 (0.01)	0.97 (0.01)	0.34 (0.07)	9.4*	—	—
Major depressive disorder	0.35 (0.05)	0.97 (0.01)	0.64 (0.06)	0.83 (0.02)	0.35 (0.05)	27.4*	0.66	0.77
Any depressive disorder	0.51 (0.05)	0.83 (0.02)	0.46 (0.04)	0.85 (0.02)	0.32 (0.05)	1.4	0.66	0.77
Generalized anxiety disorder	0.04 (0.04)	0.98 (0.1)	0.09 (0.09)	0.95 (0.01)	0.03 (0.05)	5.8*	0.72	0.81
Alcohol abuse	0.40 (0.04)	0.80 (0.02)	0.45 (0.04)	0.77 (0.02)	0.21 (0.04)	1.1	0.57	0.73
Alcohol dependence	0.60 (0.05)	0.81 (0.02)	0.46 (0.04)	0.88 (0.02)	0.37 (0.05)	8.5*	0.76	0.81
Suicide risk	0.32 (0.11)	0.87 (0.01)	0.55 (0.15)	0.97 (0.01)	0.38 (0.11)	3.6	—	—

¹Criteria for telephone and clinical psychopathology assessments are explained in Table 2 footnotes.

²True positive/True positives + False positives with clinical interview as gold standard. SEs are asymptotic.

³True negatives/True negatives + False negatives with clinical interview as gold standard. SEs are asymptotic.

⁴True positive/All positives as diagnosed on the telephone. SEs are asymptotic.

⁵True negative/All negatives as diagnosed on the telephone. SEs are asymptotic.

⁶Reliability test of extent to which telephone and clinical diagnoses agree on participant classification. SE are asymptotic.

⁷Reliability test of marginal heterogeneity to see if the telephone and clinical diagnoses used different core criteria.

* $P > 0.05$ suggesting differences between clinical interview as gold standard and telephone screening tool.

⁸Reliability test of internal consistency of the measurement items that make up the diagnosis

⁹Measure of overall accuracy based on the continuous score of a diagnostic test. It is the area under the ROC curve. AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; PTSD, post-traumatic stress disorder, ROC, receiver operating characteristic; SE, standard error.

Table 4. Sensitivity and specificity of psychopathology diagnoses by telephone within specific demographic groups

Characteristics	Major depressive disorder					
	PTSD ever in lifetime		Ever in Lifetime		Generalized anxiety disorder	
	Sensitivity (95% CI) ^a	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Sex						
Male	0.48 (0.28–0.67)	0.93 (0.91–0.96)	0.34 (0.24–0.43)	0.95 (0.92–0.97)	0.06 (0.0–0.29)	0.97 (0.96–0.99)
Female ¹	1 (0.29–1.00)	0.83 (0.70–0.92)	0.41 (0.18–0.67)	0.91 (0.78–0.97)	— ²	0.98 (0.90–1.0)
Age						
17–34	0.52 (0.30–0.74)	0.93 (0.90–0.96)	0.30 (0.19–0.41)	0.95 (0.92–0.98)	— ²	0.98 (0.96–1.0)
≥ 35	0.57 (0.18–0.90)	0.91 (0.86–0.96)	0.42 (0.28–0.58) ^a	0.92 (0.88–0.98)	0.13 (0.0–0.53)	0.97 (0.94–1.0)
Race						
White	0.55 (0.32–0.76)	0.93 (0.90–0.95)	0.34 (0.25–0.44)	0.94 (0.92–0.97)	0.05 (0.0–0.25)	0.98 (0.96–0.99)
Non-white ¹	0.50 (0.12–0.88)	0.91 (0.78–0.97)	0.38 (0.14–0.68)	0.98 (0.87–1.00)	— ^b	0.98 (0.89–1.00)

¹Exact standard errors used to calculate 95% CI due to small sample size (< 50); otherwise asymptotic standard error used based on the fact that the outcome was rare ($P < 0.05$).

² $n = 0$ for either the telephone sample or clinical interview sub-sample for the specific diagnosis.

CI, confidence interval; PTSD, post-traumatic stress disorder.

Table 5. Sensitivity and specificity of psychopathology diagnoses by telephone within specific demographic groups

Characteristics	Alcohol abuse		Alcohol dependence		Suicide	
	Sensitivity (95% CI)	Sensitivity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Sex						
Male	0.37 (0.28–0.45)	0.79 (0.74–0.83)	0.46 (0.37–0.55)	0.88 (0.85–0.92)	0.36 (0.13–0.65)	0.99 (0.98–1.00)
Female	0.67 (0.41–0.87)	0.93 (0.81–0.99)	0.43 (0.10–0.82)	0.92 (0.85–1.00)	0.20 (0.01–0.72)	0.98 (0.95–1.0)
Age						
17–34	0.45 (0.34–0.56)	0.78 (0.75–0.84)	0.60 (0.49–0.72)	0.81 (0.77–0.86)	0.20 (0.04–0.48)	0.98 (0.96–1.0)
≥ 35	0.40 (0.27–0.54)	0.75 (0.66–0.83)	0.59 (0.39–0.76)	0.83 (0.76–0.89)	0.75 (0.19–0.99)	— ¹
Race						
White	0.40 (0.31–0.47)	0.79 (0.75–0.84)	0.60 (0.49–0.70)	0.81 (0.77–0.85)	0.36 (0.13–0.65)	0.99 (0.98–1.0)
Non-White	0.50 (0.21–0.79)	0.86 (0.72–0.95)	0.60 (0.15–0.95)	0.90 (0.82–0.98)	0.20 (0.01–0.72)	0.96 (0.86–1.0)

¹n=0 for either the telephone sample or in-person clinical interview sub-sample for the specific diagnosis. CI, confidence interval.

World Health Organization Composite International Diagnostic Interview (CIDI) and the SCID for PTSD, MDD, alcohol abuse, and alcohol dependence, whereas we found differences for PTSD, MDD, GAD, and alcohol dependence.

Reliability statistics are population dependent, so it is important to understand that the findings from this military population study may not be generalized to other populations. The current study is also limited by the small percentage of women and other minorities; however, the demographics of our sample very closely mirror the overall demographics of the OHARNG. Future work should also compare the item-by-item response of the telephone assessments to the clinical interview to assess if there are internal differences within a construct.

This work suggests that the computer-assisted telephone psychopathology assessments used in the OHARNG MHI are valid and reliable research tools for the National Guard population. As compared to face-to-face interviews, the telephone assessments may also prove to be more cost-effective based on the reduced cost of travel for such a widespread population ($N = 2616$). Our telephone assessments had comparable, if slightly lower, measures of reliability as compared to the MILCO study web-based interviews (Smith *et al.*, 2007a). The web-based interviews from the MILCO study, however, resulted in a slightly lower response rate (37%) as compared to the telephone-based

assessments in this study (43%), suggesting that the CATI method of mental health assessments may prove a better tool for the OHARNG population.

Conclusion

The OHARNG MHI will continue to follow the OHARNG members over time. This longitudinal study is expected to advance the knowledge about the trajectories of post-deployment psychopathologies and facilitate enhancements in access to care and treatment of behavioral health issues among National Guard soldiers.

Acknowledgments

The research was funded by the Department of Defense Congressionally Directed Medical Research program: W81XWH0710409, the "Combat Mental Health Initiative". Results were presented at the 26th Annual Meeting of the International Society for Traumatic Stress Studies, November 4–6, 2010, Montreal, Quebec, Canada. The informatics support for this research was provided by the Michigan State University clinical and Translational Sciences Institute, through its Biomedical Research Informatics Core.

Declaration of interest statement

The authors have no competing interests.

References

- American Psychiatric Association. (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), Washington, DC, American Psychiatric Association.
- 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**(1), 75–90.
- Blanchard E.B., Jones-Alexander J., Buckley T.C., Forneris C.A. (1996) Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy*, **34**(8), 669.
- Breslau N., Kessler R.C., Chilcoat H.D., Schultz L. R., Davis G.C., Andreski P. (1998) Trauma and posttraumatic stress disorder in the community: the 1996 Detroit Area Survey of Trauma. *Archives of General Psychiatry*, **55**(7), 626–632.
- Calabrese J.R., Prescott M., Tamburrino M., Liberzon I., Slembariski R., Goldmann E., Shirley E., Fine T., Goto T., Wilson K., Ganocy S., Chan P., Serrano M.B., Sizemore J., Galea S. (2011) PTSD comorbidity and suicidal ideation associated with PTSD within the Ohio Army National Guard. *Journal of Clinical Psychiatry*, **72**(8), 1072–1078.
- Dohrenwend B., Turner J., Tursen N., Adams B., Koenen K., Marshall R. (2006) The psychological risks of Vietnam for U.S. veterans: a revisit with new data and methods. *Science*, **313**(5789), 979.
- Goldmann E., Calabrese J.R., Prescott M.R., Tamburrino M., Liberzon I., Slembariski R., Shirley E., Fine T., Goto T., Wilson K., Ganocy S., Chan P., Serrano M.B., Sizemore J., Galea S. (2012) Potentially modifiable pre-, peri-, and postdeployment characteristics associated with deployment-related posttraumatic stress disorder among Ohio Army National Guard soldiers. *Annals of Epidemiology*, **22**(49), 71–78.
- Hoge C.W., Castro C.A., Messer S.C., McGurk D., Cotting D.I., Koffman R.L. (2004) Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care [see comment]. *New England Journal of Medicine*, **351**(1), 13–22.
- Hotopf M., Hull L., Fear N.T., Browne T., Horn O., Iversen A., Jones M., Murphy D., Bland D., Earnshaw M., Greenberg N., Hughes J.H., Tate A.R., Dandeker C., Rona R., Wessely S. (2006) The health of UK military personnel who deployed to the 2003 Iraq war: a cohort study. *The Lancet*, **367**(9524), 1731.
- Iversen A., van Staden L., Hughes J., Browne T., Hull L., Hall J., Greenberg N., Rona R., Hotopf M., Wessely S., Fear N. (2009) The prevalence of common mental disorders and PTSD in the UK military: using data from a clinical interview-based study. *BMC Psychiatry*, **9**(1), 68.
- Johnson J., Maxwell A., Galea S. (2009) The epidemiology of posttraumatic stress disorder. *Psychiatric Annals*, **39**(6), 326–334.
- Kempf A.M., Remington P.L. (2007) New challenges for telephone survey research in the twenty-first century. *Annual Review of Public Health*, **28**, 113–126.
- Kessler R., Berglund P., Demler O., Jin R., Merikangas K., Walters E. (2005a) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity

- Survey Replication. *Archives of General Psychiatry*, **62**(6), 593.
- Kessler R., Chiu W., Demler O., Merikangas K., Walters E. (2005b) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, **62**(6), 617–627.
- Killgore W.D.S., Stetz M.C., Castro C.A., Hoge C.W. (2006) The effects of prior combat experience on the expression of somatic and affective symptoms in deploying soldiers. *Journal of Psychosomatic Research*, **60**(4), 379–385.
- Kroenke K., Spitzer R.L., Williams J.B. (2001) The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**(19), 606–613.
- Kulka R.A. (1990) Trauma and the Vietnam War generation: report of findings from the National Vietnam veterans readjustment study. Brunner/Mazel psychosocial stress series. 18.
- La Bash H.A.J., Vogt D.S., King L.A., King D.W. (2009) Deployment stressors of the Iraq War: insights from the mainstream media. *Journal of Interpersonal Violence*, **24**(2), 231–258.
- Milliken C.S., Auchterlonie J.L., Hoge C.W. (2007) Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. *JAMA: Journal of the American Medical Association*, **298**(18), 2141–2148.
- Remington P.L., Smith M.Y., Williamson D.F., Anda R.F., Gentry E.M., Hogelin G.C. (1988) Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–87. *Public Health Reports*, **103**(4), 366–375.
- Sheehan D.V., Lecrubier Y., Sheehan K.H., Amorim P., Janavs J., Weiller E., Hergueta T., Baker R., 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 Psychology*, **59**(Suppl 20), 22–33.
- Smith T.C., Smith B., Jacobson I.G., Corbeil T.E., Ryan M.A., Millennium Cohort Study (2007a) Reliability of standard health assessment instruments in a large, population-based cohort study. *Annals of Epidemiology*, **17**(7), 525–532.
- Smith B., Smith T., Gray G., Ryan M.A.K., Millennium Cohort Study. (2007b) When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. *American Journal of Epidemiology*, **166**(11), 1345.
- Smith T.C., Ryan M.A., Wingard D.L., Slymen D.J., Sallis J.F., Kritiz-Silverstein D., Millennium Cohort Study. (2008) New onset and persistent symptoms of post-traumatic stress disorder self reported after deployment and combat exposures: prospective population based US military cohort study. *BMJ*, **336**(7640), 366–371.
- Spitzer R.L., Kroenke K., Williams J., Löwe B. (2006) A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, **166**(73), 1092–1097.
- Terhakopian A., Sinaï N., Engel C.C., Schnurr P.P., Hoge C.W. (2008) Estimating population prevalence of posttraumatic stress disorder: an example using the PTSD checklist. *Journal of Traumatic Stress*, **21**(3), 290–300.
- Vasterling J.J., Proctor S.P., Amoroso P., Kane R., Heeren T., White R.F. (2006) Neuropsychological outcomes of army personnel following deployment to the Iraq war. *JAMA: Journal of the American Medical Association*, **296**(5), 519–529.
- Vogt D.S., Samper R.E., King D.W., King L.A., Martin J.A. (2008) Deployment stressors and posttraumatic stress symptomatology: comparing active duty and National Guard/Reserve personnel from Gulf War I. *Journal of Traumatic Stress*, **21**(1), 66–74.
- Weathers F.W., Ruscio A.M., Keane T.M. (1999) Psychometric properties of nine scoring rules for the Clinician-Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment*, **11**(2), 124–133.