J Head Trauma Rehabil
Vol. 20, No. 6, pp. 501–511

Validity of the Patient Health Questionnaire-9 in Assessing Depression Following Traumatic Brain Injury

Jesse R. Fann, MD, MPH; Charles H. Bombardier, PhD; Sureyya Dikmen, PhD; Peter Esselman, MD; Catherine A. Warms, PhD; Erika Pelzer, BS; Holly Rau, BS; Nancy Temkin, PhD

Objective: To test the validity and reliability of the Patient Health Questionnaire-9 (PHQ-9) for diagnosing major depressive disorder (MDD) among persons with traumatic brain injury (TBI). **Design:** Prospective cohort study. **Setting:** Level I trauma center. **Participants:** 135 adults within 1 year of complicated mild, moderate, or severe TBI. **Main Outcome Measures:** PHQ-9 Depression Scale, Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (SCID). **Results:** Using a screening criterion of at least 5 PHQ-9 symptoms present at least several days over the last 2 weeks (with one being depressed mood or anhedonia) maximizes sensitivity (0.93) and specificity (0.89) while providing a positive predictive value of 0.63 and a negative predictive value of 0.99 when compared to SCID diagnosis of MDD. Pearson's correlation between the PHQ-9 scores and other depression measures was 0.90 with the Hopkins Symptom Checklist depression subscale and 0.78 with the Hamilton Rating Scale for Depression. Test-retest reliability of the PHQ-9 was $r = 0.76$ and $\kappa = 0.46$ when using the optimal screening method. **Conclusions:** The PHQ-9 is a valid and reliable screening tool for detecting MDD in persons with TBI. **Key words:** *brain injury*, *depression*, *diagnosis*, *PHQ*, *reliability*, *SCID*, *screening*, *trauma*, *validity*

MAJOR depressive disorder (MDD) is reported to be the most common psychiatric disorder following traumatic brain injury (TBI).[1] Prevalence estimates have ranged widely from about 10% to more than 70%, partly due to variations in the method of ascertaining depression.[2] With greater recognition of the prevalence of MDD among people with TBI, researchers have called for improved detection and treatment of this important comorbid condition.[2–4]

Effective treatment of MDD requires accurate detection and diagnosis. Screening and diagnosis are complicated by problems with unawareness of psychological symptoms and transdiagnostic symptoms (eg, poor concentration and fatigue are associated with both TBI and MDD). Despite these problems, there is now evidence that it is scientifically and clinically appropriate to use standard *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) criteria[5] for diagnosing MDD.[1,6–9] The validity of diagnostic interviews after TBI

From the Departments of Psychiatry & Behavioral Sciences (Drs Fann and Dikmen), Rehabilitation Medicine (Drs Fann, Bombardier, Dikmen, Esselman, and Temkin and Mss Pelzer and Rau), Biobehavioral Nursing and Health Systems (Dr Warms), Neurological Surgery (Drs Dikmen and Temkin), Epidemiology (Dr Fann) and Biostatistics (Dr Temkin), University of Washington, Seattle.

This article has been supported by grant RO1 HD39415 from the National Center for Medical Rehabilitation Research. The authors thank Jason Barber, MS, for his assistance with the statistical analysis.

Corresponding author: Jesse R. Fann, MD, MPH, Department of Psychiatry & Behavioral Sciences, Box 356560, University of Washington, Seattle, WA 98195 (e-mail: fann@u.Washington.edu).

is also supported by research that shows good concordance between patient self-report and the report of independent observers on a measure that includes psychological symptoms.[10]

While the validity of diagnostic assessments is accepted, there is little consensus regarding how to screen for MDD after TBI in the clinical setting. A wide range of depression measures has been used, including the Beck Depression Inventory (BDI),[11] Center for Epidemiologic Studies Depression Scale,[12] Minnesota Multiphasic Personality Inventory,[13,14] Hopkins Symptom Checklist depression subscale (SCL-20),[15] and Neurobehavioral Functioning Inventory.[3] Few studies have tested screening measures against *DSM*-based "criterion standard" structured diagnostic interviews such as the Structured Clinical Interview for *DSM-IV* (SCID).[16] In one study that did evaluate the criterion validity of the BDI, the BDI was found to have low sensitivity (36% when the specificity was set at 80%) for major depression in people with TBI.[11]

We selected the Patient Health Questionnaire-9 (PHQ-9) depression scale as our screening measure for several reasons.[17,18] The PHQ-9 parallels the 9 diagnostic symptom criteria that define *DSM-IV* MDD. The format and temporal framework of the items also correspond to the *DSM-IV* criteria and will facilitate the follow-up review of symptoms and diagnostic process. At only 9 items, the PHQ-9 is substantially shorter than most depression screening measures. Unlike most other measures of depression, the PHQ-9 was developed, tested, and refined for use with medical patients. This is important because the criterion validity was established in a population with high rates of other physical symptoms and associated nonspecific psychological distress. This instrument has also demonstrated acceptability among nonpsychiatric patients and among busy primary care providers.[17–19]

The purpose of this study was to test the validity and reliability of the PHQ-9 in screening for MDD among persons with TBI who were enrolled in the Surveillance Phase of a larger study of the epidemiology and treatment of MDD in the year following TBI.

METHODS

Participants

From April 2001 through November 2004, patients with a TBI were recruited from Harborview Medical Center, a Level I trauma center in Seattle, Wash. Subjects were hospitalized patients who sustained a traumatic injury to the head, with either the lowest Glasgow Coma Scale (GCS) score of less than or equal to 12 or radiological evidence of acute brain abnormality. Other inclusion criteria were as follows: resident of greater Puget Sound region (King, Pierce, Kitsap, Jefferson, Mason, Thurston, or Snohomish counties); 18 years of age or older; and English speaking. Subjects were excluded on the basis of homelessness (or no contact information available), incarceration, a history of schizophrenia, or participation in an investigational drug study. The University of Washington Institutional Review Board approved all study procedures. All procedures followed Health Insurance Portability and Accountability Act guidelines.

Consecutively, eligible patients with TBI were identified via daily automatic electronic medical records queries, cross-checked against brain injury consultation lists. Research staff approached and consented eligible patients in the hospital. Patients discharged before consent were informed of the study through a letter from an attending neurosurgeon and recruited through follow-up telephone calls. Patients were required to be fully oriented on a standardized measure and to sign and return study consent forms prior to data collection.

Procedures

Participants were assessed every month for 6 months and then at 8, 10, and 12 months following injury, using a structured telephone interview conducted by trained research study assistants. Subjects who failed orientation screens were not assessed but were reevaluated monthly, resulting in some

patients entering the study 2 or more months after injury.

Subjects were screened for depression using the PHQ-9 at each telephone assessment. Depending on their score on the PHQ-9, a fraction of participants was asked to complete a SCID, preferably in person but over the telephone if the patient was unable to come in for the interview. Forty-seven percent of invited subjects completed the SCID within 7 days of taking the screening PHQ-9 and are included in our criterion validity analyses ($N = 135$), 23% completed the SCID more than 7 days after the PHQ-9, and 30% refused or missed SCID appointments. The percent of those meeting PHQ-9 screening criteria who completed the SCID within 7 days are: 10% of those with at least 5 PHQ-9 symptoms present more than half the days (suicidal ideation could be only several days), with at least one of the symptoms being a cardinal symptom (ie, anhedonia or depressed mood); 10% otherwise meeting this criteria with severity of only several days on each symptoms; 13% otherwise with a PHQ-9 sum score of ≥ 5; and 1% of those with a PHQ-9 sum score of 0–4. These completion percentages were used in weighting the analysis so that results reflect the entire screening cohort (see Statistical Analysis section).

At each SCID interview, the PHQ-9, SCL-20, and Hamilton Rating Scale for Depression (HAM-D) were also administered. The HAM-D was administered in a standardized, semistructured interview format. The PHQ-9 and SCL-20 were administered in paper-and-pencil format. If the participant was unable to read or needed assistance to complete a form, the examiner read the questions to the patient. Length of assessments varied from 30 to 90 minutes, depending on the participant's responses, speech patterns, and cognitive status.

Measures

Brain injury severity

Research study nurses reviewed medical records and coded the lowest GCS score within the first 24 hours after TBI or the first GCS score after paralytic agents were withdrawn if they were continued beyond 24 hours. Subjects were recruited into the study if their GCS score was 12 or less or if there was radiological evidence of acute brain abnormality.

Orientation

Potential participants were administered the standardized orientations scale from the Cognistat.[20] This 7-item scale covering orientation to person, age, place, and time was administered before every phone assessment until the patient was oriented. A score of at least 10 is "normal" on the basis of test norms[20] and was required before depression screening interviews took place.

Depression measures

Depression screening

The PHQ-9 depression scale was used to screen for depressive symptoms on the basis of *DSM-IV* diagnostic criteria.[17,18,21] The PHQ-9 was chosen because it has excellent internal and test-retest reliability as well as criterion and construct validity in medical samples.[17–19,21]

The PHQ-9 is a self-report measure that asks if the subject had been bothered by the following problems in the past 2 weeks: (*a*) little pleasure or interest in doing things, (*b*) feeling down, depressed, or hopeless, (*c*) sleeping too little or too much, (*d*) feeling tired or having little energy, (*e*) poor appetite or overeating, (*f*) feelings of worthlessness or guilt, (*g*) concentration problems, (*h*) psychomotor retardation or agitation, and (*i*) thoughts of suicide. Subjects were asked to rate how often each symptom occurred: 0 (not at all), 1 (several days), 2 (more than half the days), or 3 (nearly every day). It has been validated for administration over the telephone.[19,22] The PHQ-9 was generally well accepted by study participants and was administered without difficulty, typically in 2 to 10 minutes, in the vast majority of cases.

We examined several methods of depression screening using the PHQ-9. It can be

scored on the basis of at least 5 symptom endorsed "more than half the days" (suicidal ideation could be "several days"), with at least one being a "cardinal symptom," that is, either (*a*) anhedonia or (*b*) depressed mood. We examined the validity of lowering the symptom frequency threshold in the above scheme to "several days." Scores can also be based on the sum of the 9-item scores. Kroenke et al[18] suggested cut-points to identify mild (5–9), moderate (10–14), moderately severe (15–19), and severe (≥20) depression. Finally, some investigators have suggested using the cardinal symptoms of anhedonia and depressed mood (items a and b) alone as a screen for MDD.[23,24]

MDD diagnosis

The SCID was used as the criterion standard to diagnose major depression.[16] Using structured questions and a decision-tree approach, it guides clinicians through a diagnostic interview that determines the presence or absence of *DSM-IV* diagnoses. SCID criteria for a diagnosis of MDD are based on established *DSM-IV* criteria.[5] The SCID is widely considered the criterion standard method of diagnosing depression in the psychiatric literature and has been used successfully in TBI populations.[1,9,11]

The validity, reliability, and acceptability of the SCID administered by telephone versus in person have been established.[25–27] Interview questions and decision trees are directly transferable to telephone delivery without modification.

The SCID was administered by research nurse practitioners with specific training in SCID administration. The questions from the MDD module of the SCID were asked using the specific wording and established guidelines for administration and coding of SCID responses.[28] The nurses were kept unaware of PHQ-9 screening results during the period when subject with PHQ-9 less than 5 were brought in for SCIDs.

Depression symptom severity

The HAM-D is a frequently used measure of depressive symptom severity that assesses the severity of psychological and physiological symptoms of depression.[29] This study used a semistructured interview to administer the 17-item HAM-D.[30,31] The HAM-D is sensitive to change in severity of depression and has been shown to have good interrater reliability.[32] It has also established reliability, validity, and acceptability when administered over the phone.[25]

The SCL-20 is a brief self-report measure of cognitive, emotional, and somatic symptoms of depression commonly used in depression epidemiology and treatment studies. The measure, reported as a mean of the 20 items, has excellent psychometric properties and is highly sensitive to change, particularly in medical populations.[33,34]

Head injury symptoms

The Head Injury Symptom Checklist (HISC) is a list of 17 symptoms that are frequently reported in the literature as part of the sequalae of TBI (eg, headaches, dizziness, balance difficulties).[35,36] This instrument was modified to measure both frequency and bothersomeness of symptoms on 0 to 5 scales. The HISC was administered if the subject was enrolled into the antidepressant treatment study ($n = 39$).

Functional impairment

The PHQ also asks about the impact of the endorsed symptoms on their ability to do their work, take care of things at home, or get along with other people, on a scale ranging from 0 (not at all) to 3 (extremely difficult). We examined whether this interference with functioning item, using a threshold of "somewhat difficult," added to the validity of the PHQ-9 symptom items and determined the correlation of this item with the symptom items.

Health perception

The 1-item General Health Scale from the SF-36[37] was used as an indicator of overall subjective health. Subjects were asked, "In general, how would you rate your overall health?" from 1 (excellent) to 5 (poor).

Statistical analysis

Which subjects received a SCID was determined in large part by how they scored on the screening PHQ-9; because of this, we applied weights to each of the subjects inversely proportional to the rate of SCID assessments within their PHQ-9 screening criteria group. The assigned weights were applied to the sample of subjects who received both a SCID and a PHQ-9, and weighted tests of association were carried out.

For comparing depression screening methods on the basis of the PHQ-9 against the SCID, weighted calculations of sensitivity, specificity, predictive values, likelihood ratios, and kappas are reported. The effect of injury severity on sensitivity is assessed by considering only those who were diagnosed as having MDD by the SCID and forming a weighted 2 × 3 table, with rows indicating severity groups and columns indicating whether the person screened positive or negative. Each row reflects the sensitivity of the PHQ-9 within that severity group. A test for inequality of the proportions screening positive examines differential sensitivity. A similar analysis was carried out for specificity and similar analyses looked at the effect of phone versus in-person administration of the SCID. Test-retest reliability is reported as kappa with 95% confidence interval and Pearson correlation. Associations between PHQ-9 and SCL-20, HAM-D, and HISC scores are reported as Pearson correlations and t tests for difference of mean between depressed and nondepressed groups. Associations between PHQ-9 and functional impairment and health perception are reported as Spearman correlations and Mann-Whitney U tests.

RESULTS

Patient characteristics

Table 1 compares the demographic and TBI characteristics of participants who completed the PHQ-9 and the SCID versus those who did not complete a SCID. There were no significant differences between groups who did and did not complete the SCID on any of the categories.

Criterion validity

Table 2 shows the sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios, and kappas for 5 methods of depression screening using the PHQ-9. The sensitivity and specificity for these methods have been plotted, along with the receiver operator characteristic curve for the PHQ-9 sum scores (Fig 1). Using a screening criteria of at least 5 PHQ-9 symptoms present at least several days over the last 2 weeks, with at least one of the symptoms being a cardinal symptom (ie, anhedonia or depressed mood), was found to be the optimal screening criterion, maximizing sensitivity (0.93) and specificity (0.89), while providing a positive predictive value of 0.63, a negative predictive value of 0.99, a positive likelihood ratio of 8.58, a negative likelihood ratio of 0.08, and a κ of 0.69. Using a PHQ-9 sum score cutoff of 12 or more also provided good sensitivity (0.85) and specificity (0.94). The area under the PHQ-9 sum score ROC curve is 0.97, suggesting a test that discriminates well between persons with and without major depression.

We also examined how using the interference with functioning question on the PHQ-9 impacted its sensitivity and specificity. We found that requiring endorsement of at least "somewhat difficult" functioning as a result of the endorsed depressive symptoms decreased the sensitivity to 0.81, but increased the specificity to 0.93 when using the optimal screening method of 5 or more symptoms present at least several days, with at least one being a cardinal symptom.

Construct validity

Convergent validity

The relationships between the PHQ-9 and other measures of depression were examined to determine convergent validity. The Pearson's correlation between the PHQ-9 score with other depression measures were 0.90 ($P < .001$) with the SCL-20 and 0.78 ($P <$

Table 1. Patient and injury characteristics*

	Total sample ($N = 478$)	SCID never administered ($n = 343$)	SCID administered ($n = 135$)	*P*
Mean age (SD)	42 (17.9)	43 (17.4)	42 (16.8)	NS
Male gender	339 (70.9%)	245 (71.4%)	94 (69.6%)	NS
Race				NS
Caucasian	431 (90.2%)	311 (90.7%)	120 (88.9%)	
African American	20 (4.2%)	13 (3.8%)	7 (5.2%)	
Asian/Pacific Islander	15 (3.1%)	10 (2.9%)	5 (3.7%)	
Other	12 (2.5%)	9 (2.6%)	3 (2.2%)	
Ethnicity				NS
Hispanic	19 (4.0%)	16 (4.7%)	3 (2.2%)	
Non-Hispanic	459 (96.0%)	327 (95.3%)	132 (97.8%)	
Married	178 (37.2%)	130 (38.0%)	48 (35.6%)	NS
High school or greater (includes GED)	430 (90.5%)†	308 (90.3%)‡	122 (91.0%)§	NS
Mechanism of injury				NS
Vehicular accident	223 (46.7%)	167 (48.7%)	56 (41.5%)	
Fall	155 (32.4%)	110 (32.1%)	45 (33.3%)	
Assault (penetrating or blunt)	46 (9.6%)	31 (9.0%)	15 (11.1%)	
Recreational/sports	26 (5.4%)	19 (5.5%)	7 (5.2%)	
Other	28 (5.9%)	16 (4.7%)	12 (8.9%)	
Glasgow Coma Scale				NS
Complicated mild (13–15)	254 (53.1%)	178 (51.9%)	76 (56.3%)	
Moderate (9–12)	111 (23.2%)	86 (25.1%)	25 (18.5%)	
Severe (≤8)	113 (23.6%)	79 (23.0%)	34 (25.2%)	
Mean months since traumatic brain injury (SD)			3.8 (2.8)	

*SCID indicates Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders* (4th ed).
†$N = 475$.
‡$n = 341$.
§$n = 134$.

.001) with the 17-item HAM-D for the 198 subjects with all 3 measures completed on the same day. The optimal method of screening for MDD on the basis of 5 or more symptoms present at least several days (score of ≥1), with at least one cardinal symptom, was significantly associated with a higher SCL-20 (difference = 1.19; $t = 13.82$, $df = 196$, $P < .001$) and a higher 17-item HAM-D (difference = 10.2; $t = 10.98$, $df = 196$, $P < .001$).

Convergent validity was also examined in relation to functional impairment and general health. The PHQ-9 score was correlated with the functional impairment item (Pearson's coefficient = 0.59, $P < .001$) and general health perception (Pearson's coefficient = 0.40, $P < .001$). A determination of MDD based on 5 or more symptoms present at least several days (score of ≥1), with at least one cardinal symptom, was significantly associated with functional impairment ($Z = 10.6$, $P < .001$) and general health perception ($Z = 5.8$, $P < .001$).

Discriminant validity

The PHQ-9's correlation with the HISC was examined in the 39 subjects who completed both measures after consenting to the Treatment Phase of the parent study. We excluded

Table 2. Criterion validity ($N = 135$)*

PHQ-9 screening criteria	% Meeting PHQ-9 screening criteria	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Likelihood ratio--positive	Likelihood ratio--negative	κ
Total PHQ-9 score $\geq$ 10	22.5	0.88 (0.66, 0.98)	0.90 (0.83, 0.95)	0.63 (0.44, 0.80)	0.97 (0.92, 1.00)	8.77	0.14	0.67
Total PHQ-9 score $\geq$ 10, at least one cardinal symptom scored $\geq$2	16.3	0.73 (0.50, 0.89)	0.95 (0.89, 0.98)	0.72 (0.49, 0.89)	0.95 (0.89, 0.98)	13.23	0.29	0.67
At least 5 symptoms scored $\geq$ 2, at least one cardinal symptom†	11.6	0.60 (0.37, 0.80)	0.98 (0.93, 1.00)	0.84 (0.57, 0.97)	0.93 (0.86, 0.97)	26.81	0.41	0.65
At least 5 symptoms scored $\geq$ 1, at least one cardinal symptom	24.1	0.93 (0.74, 1.00)	0.89 (0.82, 0.94)	0.63 (0.44, 0.79)	0.99 (0.94, 1.00)	8.58	0.08	0.69
At least one cardinal symptom scored $\geq$ 2	20.5	0.75 (0.52, 0.91)	0.90 (0.83, 0.95)	0.59 (0.39, 0.78)	0.95 (0.89, 0.98)	7.44	0.28	0.59

*PHQ-9 indicates Patient Health Questionnaire-9; CI, confidence internal.
†Suicidal ideation item $\geq$ 1.

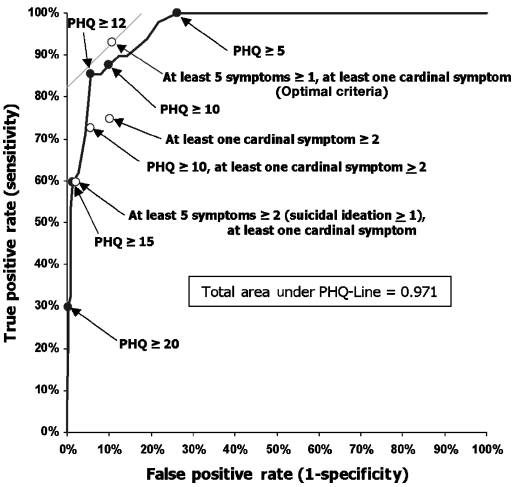

Figure 1. PHQ versus SCID: receiver operator characteristic curve. PHQ indicates Patient Health Questionnaire; SCID, Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders*, (4th ed).

HISC items overlapping with *DSM-IV* MDD symptoms, that is, fatigue, concentration, and sleep. Pearson's correlations were 0.49 ($P = .002$) for bothersomeness of symptoms and 0.44 ($P = .005$) for frequency of symptoms. Among these subjects, the PHQ-9, HISC, SCL-20, and HAM-D were all completed in-person on the same day. The correlation of the PHQ-9 with the SCL-20 was 0.84 ($P < .001$) and the HAM-D was 0.67 ($P < .001$) in this subgroup of subjects.

Reliability

Test-retest reliability of the PHQ-9 was calculated for the 132 assessments repeated within 7 or fewer days. The Pearson's correlation was 0.76 ($P < .001$) for the total score and the κ was 0.46 (95% CI: 0.31, 0.61; $P < .001$) for MDD as determined by 5 or more symptoms present at least several days (score of ≥ 1), with at least one cardinal symptom present.

Validity modifiers

We found that neither TBI severity nor whether the SCID was performed in person or over the telephone significantly modified the criterion validity (sensitivity and specificity) of the PHQ-9.

DISCUSSION

Our study found that the PHQ-9 depression scale administered by telephone or by paper and pencil by the patient is a valid and reliable screening tool for major depression in oriented persons with TBI. It was generally well tolerated, brief, and simple to administer in this diverse head-injured patient population.

On the basis of sensitivity and specificity, we found that the presence of 5 or more depressive symptoms for at least several days over the last 2 weeks (score of ≥ 1 on the PHQ-9), with at least one symptom being a cardinal symptom, that is, anhedonia or depressed mood, was the optimal screening method for MDD, as defined by the sum of sensitivity and specificity. Using this cutoff, one misses very few depressed patients (NPV = 0.99), although about 3 in 8 cases who screen positive would not be diagnosed as having MDD if the SCID were done (PPV = 0.63). However, the negative predictive value of 0.99 (meaning that a person with a negative depression screen had a .99 probability of not having MDD) and negative likelihood ratio of 0.08 (meaning that a negative depression screen was less than 0.1 times as likely to be seen in someone with MDD than in someone without MDD) are favorable for depression screening purposes. Using the interference with functioning question, which corresponds with MDD Criterion C in *DSM-IV*, compromised sensitivity while improving specificity only slightly; therefore, this question is unlikely to be useful for depression screening unless maximizing specificity at the potential expense of sensitivity is required.

The PHQ-9 had excellent convergent validity with 2 commonly used depression measures, the self-report SCL-20 and the clinician-rated HAM-D. It also correlated with functioning and quality of life, domains known to be affected by depression.[6,8,38] The 0.49 correlation of the PHQ-9 with the HISC, as expected,

was lower than with the depression measures. However, the presence of a certain level of correlation of depressive symptoms with head injury symptoms is not surprising, given the finding that depression can amplify somatic symptoms in patients with TBI.[6]

Using a PHQ-9 sum score cutoff of 12 provided the best MDD screening criterion on the basis of a PHQ-9 sum score. Using the standard PHQ-9 scoring—that is, at least 5 symptoms present at least half the days (suicidal ideation could be present only several days), with at least one being a cardinal symptom—provided the best specificity and positive likelihood ratio, but led to poor sensitivity. The 2-item screening method that uses the first 2 *DSM-IV* cardinal symptoms showed acceptable specificity, but poor sensitivity. The method ultimately used in clinical and research practice will depend on the specific needs and estimated prevalence of depression in the population in question.

Our finding that lowering the symptom duration criteria to "several days" while still requiring 5 symptoms and at least one cardinal symptom showed the best overall sensitivity and specificity warrants further discussion. While persons with TBI may be able to accurately report the presence of depressive symptoms, ongoing memory impairment may lead to difficulty reporting duration of depressive symptoms without some cueing or prompting, such as in a more detailed clinical assessment. Moreover, this relatively young, predominantly male TBI population may tend to underreport depressive symptoms.[39–41] As a result, lowering the duration criteria from *DSM-IV*'s "nearly every day"[5] and the PHQ-9's "more than half the days"[17] to "several days" may be needed in TBI populations for purposes of screening for probable major depression.

Test-retest reliability of the PHQ-9 was fair and may have been compromised by fluctuation in depression symptoms within the up to 7-day window between interviews. Furthermore, differences in telephone and paper-and-pencil administration of the PHQ-9 may have influenced retest reliability. We found, however, that TBI severity and in-person versus telephone SCID administration did not modify the PHQ-9's criterion validity.

To our knowledge, this is the first study that demonstrates acceptable reliability and validity of a brief screening measure for major depression measured by the SCID in persons with TBI. The psychometric properties of the PHQ-9 in this population compare favorably with the diagnostic accuracy of depression screening measures in other medical populations, where the median sensitivity and specificity of screening measures is 85% and 74%, respectively.[42] As with use of any depression screening tool, a follow-up detailed clinical assessment to make a definitive diagnosis of MDD is indicated for those who meet initial screening criteria.

The PHQ-9 has several potential advantages as a depression screening tool in TBI populations. Items on the PHQ-9 correspond directly with *DSM-IV* Criterion A (and Criterion C if one also uses the interference with functioning question) used to determine MDD.[5,17,18] Also, this measure is considerably shorter than other commonly used depression screening measures, such as the BDI,[43] the Center for Epidemiologic Studies Depression Scale,[44] and the SCL-20,[33] and can be administered by self-report or by nonclinician interview.[19] Finally, the results of this measure yield both the possible diagnosis of major depression as well as an assessment of symptom severity, 2 useful elements in detecting MDD and monitoring course and treatment response. The PHQ-9 has shown promise as a valid measure of depression treatment response in some medical populations[22]; however, further research on its predictive validity is necessary in TBI patients.

Limitations of this study include relatively small sample size and changes in inclusion criteria for performing SCID assessments that required weighting of the data. Because this study was a part of a larger study that sought to identify and enroll patients with MDD into an antidepressant treatment trial and was not designed solely to test the validity of the PHQ-9 and SCID administered by phone, there was

a sampling bias of patients with higher PHQ-9 scores to be brought in for in-person SCIDs. We attempted to address this bias and improve the generalizability of our results by recruiting a subset of patients with PHQ-9 total scores in the lowest range (0–4) and by using weighted variables for the primary analyses. The study is also limited by the fact the nurses who administered the SCID were not kept completely unaware of screening test results, potentially introducing a measurement bias. However, we did employ masking procedures during the period when we recruited patients for SCID assessments who had PHQ-9 scores less than 5.

Our findings are encouraging in that the PHQ-9 depression scale appears to be a feasible, reliable, and valid screening tool for major depression, a common disorder that significantly contributes to morbidity, in persons who have sustained a wide range of TBI severities. Further research will be necessary to cross-validate our findings and examine if the PHQ-9 is sensitive to change over time, such as with treatment, in this population. The PHQ-9 may be also useful in larger studies of the epidemiology of depression in TBI populations and to compare the epidemiology and phenomenology of depression in TBI populations with other medical and trauma populations.

REFERENCES

1. Hibbard MR, Uysal S, Kepler K, Bogdany J, Silver J. Axis I psychopathology in individuals with traumatic brain injury. *J Head Trauma Rehabil.* 1998;13:24–39.
2. Rosenthal M, Christensen BK, Ross TP. Depression following traumatic brain injury. *Arch Phys Med Rehabil.* 1998;79:90–103.
3. Seel RT, Kreutzer JS, Rosenthal M, Hammond FM, Corrigan JD, Black K. Depression after traumatic brain injury: a National Institute on Disability and Rehabilitation Research Model Systems multicenter investigation. *Arch Phys Med Rehabil.* 2003;84:177–184.
4. Kreutzer JS, Seel RT, Gourley E. The prevalence and symptom rates of depression after traumatic brain injury: a comprehensive examination. *Brain Inj.* 2001;15:563–576.
5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 4th ed. Washington, DC: American Psychiatric Association; 2000.
6. Fann JR, Katon WJ, Uomoto JM, Esselman PC. Psychiatric disorders and functional disability in outpatients with traumatic brain injuries. *Am J Psychiatry.* 1995;152:1493–1499.
7. Jorge RE, Robinson RG, Arndt S. Are there symptoms that are specific for depressed mood in patients with traumatic brain injury? *J Nerv Ment Dis.* 1993;181:91–99.
8. Jorge RE, Robinson RG, Moser D, Tateno A, Crespo-Facorro B, Arndt S. Major depression following traumatic brain injury. *Arch Gen Psychiatry.* 2004;61:42–50.
9. Rapoport MJ, McCullagh S, Shammi P, Feinstein A. Cognitive impairment associated with major depression following mild and moderate traumatic brain injury. *J Neuropsychiatry Clin Neurosci.* 2005;17:61–65.
10. Seel RT, Kreutzer JS, Sander AM. Concordance of patients' and family members' ratings of neurobehavioral functioning after traumatic brain injury. *Arch Phys Med Rehabil.* 1997;78:1254–1259.
11. Sliwinski M, Gordon WA, Bogdany J. The Beck Depression Inventory: is it a suitable measure of depression for individuals with traumatic brain injury? *J Head Trauma Rehabil.* 1998;13:40–46.
12. Dikmen SS, Bombardier CH, Machamer ME, Fann JR, Temkin NR. Natural history of depression in traumatic brain injury. *Arch Phys Med Rehabil.* 2004;85:1457–1465.
13. Burton L, Volpe B. Sex differences in emotional status of traumatically brain-injured patients. *J Neurorehabil.* 1988;2:151–157.
14. Dikmen S, Reitan RM. Emotional sequelae of head injury. *Ann Neurol.* 1977;2:492–494.
15. Linn RT, Allen K, Willer BS. Affective symptoms in the chronic stage of traumatic brain injury: a study of married couples. *Brain Inj.* 1994;8:135–147.
16. First MD, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). Clinical Version*. Washington, DC: American Psychiatric Press; 1996.
17. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA.* 1999;282:1737–1744.
18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606–613.
19. Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report

instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affect Disord.* 2004;81:61–66.
20. Kiernan RJ, Mueller J, Langston JW, Van Dyke C. The neurobehavioral cognitive status examination: a brief but quantitative approach to cognitive assessment. *Ann Intern Med.* 1987;107:481–485.
21. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am J Obstet Gynecol.* 2000;183:759–769.
22. Lowe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. *Med Care.* 2004;42:1194–1201.
23. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med.* 1997;12:439–445.
24. Henkel V, Mergl R, Coyne JC, Kohnen R, Moller HJ, Hegerl U. Screening for depression in primary care: will one or two items suffice? *Eur Arch Psychiatry Clin Neurosci.* 2004;254:215–223.
25. Simon GE, Revicki D, VonKorff M. Telephone assessment of depression severity. *J Psychiatr Res.* 1993;27:247–252.
26. Cacciola JS, Alterman AI, Rutherford MJ, McKay JR, May DJ. Comparability of telephone and In-person structured clinical interview for DSM-III-R (SCID) diagnoses. *Assessment.* 1999;6:235–242.
27. Allen K, Cull A, Sharpe M. Diagnosing major depression in medical outpatients: acceptability of telephone interviews. *J Psychosom Res.* 2003;55:385–387.
28. First M, Gibbon M, Spitzer R, Williams J. *User's Guide for the Structured Clinical Interview for DSM IV Axis I Disorders.* New York: Biometrics Research Department, New York State Psychiatric Institute; 1996.
29. Hamilton M. Development of a rating scale of primary depressive illness. *Br J Soc Clin Psychol.* 1967;6:278–296.
30. Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry.* 1988;45:742–747.
31. Williams JB. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci.* 2001;251(suppl 2):II6–II12.
32. Potts MK, Daniels M, Burnam MA, Wells KB. A structured interview version of the Hamilton Depression Rating Scale: evidence of reliability and versatility of administration. *J Psychiatr Res.* 1990;24:335–350.
33. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale—preliminary report. *Psychopharmacol Bull.* 1973;9:13–28.
34. Derogatis LR. *SCL-90-R Administration, Scoring, and Procedures Manual*. 3rd ed. Minneapolis, Minn: National Computer Systems Inc; 1994.
35. Dikmen S, McLean A, Temkin N. Neuropsychological and psychosocial consequences of minor head injury. *J Neurol Neurosurg Psychiatry.* 1986;49:1227–1232.
36. McLean A Jr, Dikmen SS, Temkin NR. Psychosocial recovery after head injury. *Arch Phys Med Rehabil.* 1993;74:1041–1046.
37. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36), I: conceptual framework and item selection. *Med Care.* 1992;30:473–483.
38. Christensen B, Ross T, Kotasek R, Rosenthal M, Henry R. The role of depression in rehabilitation outcomes in the acute recovery of patients with TBI. *Adv Med Psychother.* 1994;7:23–38.
39. Stommel M, Given BA, Given CW, Kalaian HA, Schulz R, McCorkle R. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry Res.* 1993;49:239–250.
40. Novy DM, Nelson DV, Averill PM, Berry LA. Gender differences in the expression of depressive symptoms among chronic pain patients. *Clin J Pain.* 1996;12:23–29.
41. Salokangas RK, Vaahtera K, Pacriev S, Sohlman B, Lehtinen V. Gender differences in depressive symptoms. An artefact caused by measurement instruments? *J Affect Disord.* 2002;68:215–220.
42. Williams JW Jr, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry.* 2002;24:225–237.
43. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961;4:561–571.
44. Radloff L. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977;1:385–401.