

Do Somatic and Cognitive Symptoms of Traumatic Brain Injury Confound Depression Screening?

Karon F. Cook, PhD, Charles H. Bombardier, PhD, Alyssa M. Bamer, MPH, Seung W. Choi, PhD, Kurt Kroenke, MD, Jesse R. Fann, MD, MPH

ABSTRACT. Cook KF, Bombardier CH, Bamer AM, Choi SW, Kroenke K, Fann JR. Do somatic and cognitive symptoms of traumatic brain injury confound depression screening? *Arch Phys Med Rehabil* 2011;92:818-23.

Objective: To evaluate whether items of the Patient Health Questionnaire 9 (PHQ-9) function differently in persons with traumatic brain injury (TBI) than in persons from a primary care sample.

Design: This study was a retrospective analysis of responses to the PHQ-9 collected in 2 previous studies. Responses to the PHQ-9 were modeled using item response theory, and the presence of DIF was evaluated using ordinal logistic regression.

Setting: Eight primary care sites and a single trauma center in Washington state.

Participants: Participants (N=3365) were persons from 8 primary care sites (n=3000) and a consecutive sample of persons with complicated mild to severe TBI from a trauma center who were 1 year postinjury (n=365).

Interventions: Not applicable.

Main Outcome Measure: PHQ-9.

Results: No PHQ-9 item demonstrated statistically significant or meaningful DIF attributable to TBI. A sensitivity analysis failed to show that the cumulative effects of nonsignificant DIF resulted in a systematic inflation of PHQ-9 total scores. Therefore, the results also do not support the hypothesis that cumulative DIF for PHQ-9 items spuriously inflates the numbers of persons with TBI screened as potentially having major depressive disorder.

Conclusions: The PHQ-9 is a valid screener of major depressive disorder in people with complicated mild to severe TBI, and all symptoms can be counted toward the diagnosis of major depressive disorder without special concern about overdiagnosis or unnecessary treatment.

Key Words: Brain injuries; Depression; Diagnosis; Psychometrics; Rehabilitation.

© 2011 by the American Congress of Rehabilitation Medicine

MAJOR DEPRESSIVE DISORDER is a highly prevalent and disabling complication of TBI. The prevalence of MDD has been estimated as high as 53% during the first year after TBI.¹ Up to 62% of TBI survivors develop MDD within the first 5 to 8 years after injury.² Psychiatric disorders, especially depression, may be among the most disabling consequences of TBI.³ Depression is more closely associated with persistence and recurrence of TBI-related functional disability than initial injury severity or persisting cognitive impairment.⁴ Effective identification and treatment of MDD after TBI have the potential to prevent or reverse significant disability in this population.⁵ However, the professional time and mental health expertise needed to conduct full psychiatric diagnostic assessments are often lacking in acute and postacute TBI care. Therefore, simple brief screening instruments are needed to identify people suspected of having MDD so that they can be evaluated and treated as needed.

A major impediment to improved screening in TBI is the longstanding and widespread belief that depression measures may not be valid because they include items that could be indicators of more than 1 diagnosis (ie, transdiagnostic symptoms).⁶ In people with TBI, fatigue, poor concentration, and sleep disturbance have been singled out as symptoms that are potentially invalid indicators of MDD.⁷ These symptoms could be attributable to TBI rather than to MDD.^{6,8-11} In the absence of definitive studies on this topic, clinicians may ignore these symptoms to avoid overdiagnosis. If these symptoms are valid indicators of MDD, this practice could lead to significant underdetection and undertreatment.

Studies are needed that can help clinicians and researchers discern whether cognitive and somatic symptoms can be counted toward a diagnosis of MDD in people with TBI versus ignored or subjected to a correction factor.⁸ Such studies are challenging because depression measures are generally scored

From the Departments of Rehabilitation Medicine (Cook, Bombardier, Bamer) and Psychiatry and Behavioral Sciences (Fann), University of Washington School of Medicine, Seattle, WA; Feinberg School of Medicine, Northwestern University, Chicago, IL (Choi); Indiana University School of Medicine, VA Center of Excellence for Implementing Evidence-Based Practice, and Regenstrief Institute, Indianapolis, IN (Kroenke).

Summary of results presented to the International Society for Quality of Life Research, September 27-30, 2010, London, United Kingdom.

Supported by the Department of Education (National Institute on Disability and Rehabilitation Research) (grant no. H133B080024, and by the National Center for Medical Rehabilitation Research, the National Institute of Child Health and Human Development, and National Institutes of Health (grant no. R01 HD39415).

The contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the U.S. government.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Correspondence to Karon F. Cook, PhD, University of Washington Center on Rehabilitation Research, 4907 25th Ave NE, Seattle, WA 98105, e-mail: karonc2@u.washington.edu. Reprints are not available from the author.

0003-9993/11/9205-0078\$36.00/0
doi:10.1016/j.apmr.2010.12.008

List of Abbreviations

CFA	confirmatory factor analysis
DIF	differential item functioning
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders 4th edition
EFA	exploratory factor analysis
GCS	Glasgow Coma Scale
GRM	graded response model
IRT	item response theory
MDD	major depressive disorder
PHQ-9	Patient Health Questionnaire 9
TBI	traumatic brain injury

using classic test theory in which item scores are manipulated (eg, summing across all items) to obtain a total score. Between-group differences in total scores may be a result of actual differences in symptom levels or differences in how participants in different patient subgroups respond to items. In recent years, modern psychometric methods such as IRT and evaluations of DIF have been applied to the measurement of health outcomes.^{12,13} These methods are well suited to explore the impact of transdiagnostic symptoms on screening for MDD in populations with TBI. If how people respond to an item is influenced not only by their level of depression but also by diagnostic group (eg, primary care vs TBI), then the item has DIF with respect to diagnosis. Thus, assessing DIF is an elegant and targeted approach to evaluating the validity of items as indicators of depression across different populations.

In the study reported here, we compared responses to PHQ-9 items among people with TBI versus primary care patients. We tested the hypothesis that the PHQ-9 items have meaningful levels of DIF. We predicted that items assessing fatigue, poor concentration, and sleep disturbance would be significant sources of DIF.⁷ We also predicted that the cumulative effects of DIF would result in inflated PHQ-9 total scores among people with TBI compared with people in primary care.

METHODS

Instrumentation

The items of the PHQ-9 are presented in Appendix 1. We chose to study the PHQ-9 depression scale¹⁴ for several reasons. First, unlike other depression screening measures, it does not contain extraneous items but completely parallels the 9 symptoms used to diagnose MDD according to DSM-IV criteria. Next, the PHQ-9 is short, easy to administer, and simple to score. It consists of 9 items with response options of 0 to 3 (not at all, several days, more than half the days, almost every day) and has a summative score range of 0 to 27. In a primary care setting, a cutoff score of 10 or more resulted in good sensitivity (.88) and specificity (.88) in identifying MDD compared with a structured diagnostic interview.¹⁵ The PHQ-9 has become one of the most widely used depression measures in numerous types of clinical populations and settings.¹⁶ Finally, prior evidence for the validity of the PHQ-9 in people with TBI has been gathered. Using a scoring procedure that parallels the DSM-IV paradigm (requiring endorsement of 5 or more symptoms, at least 1 of which must be depressed mood or anhedonia), Fann et al¹⁷ reported a sensitivity of 93% and a specificity of 89% compared with a diagnosis of MDD based on the criterion standard diagnostic interview, the Structured Clinical Interview for DSM-IV. Nevertheless, the question remains whether cognitive and somatic items should be counted toward a diagnosis of MDD in people with TBI because, whether the symptom is elicited via questionnaire or structured interview, transdiagnostic symptoms may spuriously inflate estimates of depression prevalence in people with TBI.

Sample

The current study is a retrospective analysis of data collected in 2 previous studies. The institutional review boards of each study approved the study protocols. Spitzer et al¹⁸ recruited a sample of 3000 persons from 8 primary care sites to evaluate the validity of the self-administered Primary Care Evaluation of Mental Disorders Patient Health Questionnaire for diagnosing mental disorders in primary care. Potential participants were invited if they were 18 years or older at the time of recruitment. Of the 3000 participants, 1578 were recruited from

family practices and 1422 from general internal medicine practices. To minimize sampling bias, 2 recruitment strategies were used: recruitment of consecutive patients, or recruitment of every *n*th patient until the recruitment quota was reached. All patients completed PHQ-9 items before seeing their physicians. The TBI sample was from a cohort study that followed persons for 1 year after injury. Data were collected via structured telephone interviews at months 1 through 6 and at 8, 10, and 12 months.¹ Data for the current study were obtained at 12 months and included 365 participants (79% of eligible cases). Eligibility criteria included admission to Harborview Medical Center (a level I trauma center in Seattle, WA) with TBI and radiologic evidence of acute, traumatically induced brain abnormality or GCS score lower than 13 (based on the lowest score within 24 hours after admission or the first after paralytic agents were withdrawn). Participants were at least 18 years old and English-speaking. Excluded were those with uncomplicated mild TBI (GCS 13–15 and no radiologic abnormality). Other exclusion criteria were homelessness, no contact information, incarceration, and schizophrenia.

Analyses

IRT assumption of unidimensionality. Using standard psychometric practice,¹⁹ we evaluated the dimensionality of the data using both EFA and CFA. EFAs (with promax rotation) and CFAs (hypothesizing a unidimensional model) were conducted separately in the primary care and TBI samples. MPlus software^a was used for these analyses.

DIF analyses. We used the software package LORDIF^b to evaluate DIF. LORDIF assesses DIF using an ordinal logistic regression framework.²¹ Analysis of DIF occurs at the item level, where the dependent variable is the item response category selected. In the case of the PHQ-9, values of the dependent variable range from 0 to 3, corresponding to the 4 possible responses to PHQ-9 items. A base model (model 1) is posited in which only level of depression predicts people's responses. A second model (model 2) is posited in which both level of trait and group membership (primary care vs TBI) predict responses to an item. Statistically significant DIF is indicated if, based on chi-square statistic values, model 2 is significantly better than model 1 at predicting item response. A third model (model 3) includes an interaction term and tests whether DIF has a consistent impact across levels of depression symptomatology (uniform DIF) or whether the impact of DIF varies by symptom level (nonuniform DIF).

Identification of DIF based on statistical criteria alone is problematic because chi-square values are heavily influenced by sample size.²² Therefore, standards have been proffered for identifying "meaningful DIF." Zumbo²³ suggested that a change in pseudo-*R*² statistic less than .13 represents negligible DIF. Others have offered standards based on percent change in standardized regression coefficient, beta, obtained when comparing model 1 (no group effect) with model 2 (group effect). The suggested criteria for meaningful DIF range from 10% change in beta²⁴ to 5% change in beta.²⁵ In the current study, we compared findings with each of these criteria used for meaningful DIF—change in pseudo-*R*² statistic less than .13, 5% change in beta, and 10% change in beta.

Correction for DIF. To evaluate the impact of DIF, it is necessary to quantify how much difference it makes in a person's scores. This is accomplished by using the LORDIF software. Any item found to have DIF is recalibrated to the GRM, the IRT model used in the DIF calibration²⁰; that is, the item's parameters are re-estimated separately for the primary care and for the TBI samples, yielding group-specific item parameters. Scores are then estimated based on this recalibra-

tion, resulting in “DIF-corrected” scores. The difference between scores before and after DIF correction quantifies the impact of DIF on PHQ-9 scores. If, for example, scores for persons with TBI were much higher before than after DIF correction, this would suggest that the PHQ-9 overestimates the level of depression in TBI populations.

Sensitivity analysis. As stated, there are standards for what constitutes meaningful DIF. Results of DIF analyses could be affected by choice of a criterion. Further, it is possible that no single item would reach standard criteria for meaningful DIF, but several items might have negligible amounts of DIF. If the direction of DIF for these items was the same (eg, all items overestimating depression in TBI), the cumulative effect might be meaningful even if the impact on individual items was negligible. Another possibility is that items would have negligible DIF, but the impact of that DIF would be concentrated at particular levels of depression. For example, if the impact of DIF were concentrated around threshold scores (eg, a summative score of 10, a proposed threshold for identifying moderate depression severity), this would have implications for the usefulness of the PHQ-9 as a screening instrument for persons with TBI.

To test these possibilities, we conducted a sensitivity analysis in which we set a sensitive chi-square criterion ($P < 0.1$) for DIF to ensure that we identified items with negligible DIF. We then corrected scores for this overidentified set of DIF items as described and estimated the cumulative impact on scores for different levels of depression.

Because our TBI sample was largely men, we completed a follow-up analysis in which we divided the combined sample (TBI and primary care) by sex. We then used our 3 DIF criteria to evaluate whether there were differences in PHQ-9 item functions by sex.

RESULTS

Sample Characteristics

The primary care and TBI samples have been fully described elsewhere.^{1,18} Table 1 presents a comparative summary of the demographics of the samples. The samples were similar by age. The smaller percentage of women in the TBI sample reflects the demographics of this condition.²⁶ PHQ-9 scores for the primary care sample ranged from 9 to 36 (mean \pm SD, 14.0 \pm 5.8). In the TBI sample, scores ranged from 9 to 35 (mean \pm SD, 13.6 \pm 5.8).

Table 1: Sample Characteristics

Demographics	Primary Care (n=3000)	TBI (n=365)
Age, mean \pm SD	46 \pm 17.2	43 \pm 17.7
Race/ethnicity* (%)		
African American or black	12.2	5.8
Hispanic	3.8	2.5
White (not Hispanic)	74.0	87.1
Asian/Pacific islander	1.4	2.0
American Indian/Alaskan native	0.0	1.4
Other or endorsed more than 1 category	2.3	1.4
Missing	6.2	0.0
Sex (%)		
Women	66	28

*Percentages do not total 100% because of rounding error.

Analyses

IRT assumption of unidimensionality. The EFA results indicated similar factor structures for the primary care and TBI data. First factor loadings were similar across items, ranging from .73 to .88 in the primary care sample and .64 to .89 in the TBI sample. The first factor accounted for 68.7% of the common variance in the primary care sample and 64.7% in the TBI sample. The ratios between first and second eigenvalues in primary care and TBI samples were 9.8 and 6.5, respectively. These results support the unidimensionality of the data.¹⁹ CFAs were conducted in the 2 samples, and fit statistics were generated. Values for the primary care and TBI results were, respectively, as follows: comparative fit index, .97 and .98; root mean square error of approximation, .08 and .08; and standardized root mean square residual, .04 and .06. These results also support the unidimensionality of the data.

DIF analyses. Based on the 3 standard criteria selected, none of the PHQ-9 items had meaningful DIF. In fact, even under much more sensitive criteria (ie, change in pseudo- R^2 statistic $< .01$; 2.5% change in beta), no items were identified with meaningful DIF.

Sensitivity analysis. To explore the potential cumulative impact of the presence of multiple items with small amounts of DIF, we ran a DIF exploration with a sensitive chi-square criterion ($P < .11$). This approach “tricked” the LORDIF software into identifying items with negligible DIF and generating scores corrected for the cumulative impact of these items. Based on this sensitive chi-square criterion, 4 PHQ-9 items were identified—items 4, 5, 7, and 8 (see Appendix 1). We emphasize that the levels of DIF in these items did not meet criteria for meaningful DIF. Therefore, it would be inappropriate to generate hypotheses regarding why these 4 items were tagged based on this overly sensitive chi-square criterion. The purpose of this analysis was to evaluate whether the cumulative effect of negligible DIF had a meaningful impact on scores.

To answer this question, we compared scores with and without DIF correction for persons with TBI only. The results are presented in figure 1. With IRT models, scores are calibrated in logit units that range from approximately -4 to $+4$. However, logit scores can be associated with their raw score equivalents. To assist in interpretation of results, lines have been added to figure 1 that correspond to PHQ-9 raw scores of 5, 10, 15, and 20, representing cutpoints for mild, moderate, moderately severe, and severe levels of depression, respectively.¹⁴ As figure 1 indicates, the scores were highly correlated ($r = .9998$), and correcting for DIF had very little impact on scores. An identity line is included in figure 1. Data points above the line are cases in which the uncorrected scores underestimated levels of depression; those below the line are cases in which the uncorrected scores overestimated levels of depression. The difference between mean scores corrected and not corrected for DIF was minimal, approximately .01 logit units or the equivalent of approximately .05 in raw score units (scale, 0–27).

Another way to evaluate the impact of DIF is at the person level. This was accomplished by first identifying the number of persons with TBI who would be classified as having minimal, mild, moderate, moderately severe, and severe depression based on uncorrected PHQ-9 scores. We then identified the GRM-calibrated scores (not corrected for DIF) that would result in the same classifications as the raw score cutpoints. The equivalent thresholds are reported in table 2. Next we applied these GRM cutpoints to the DIF-corrected scores. Table 2 reports the number of persons classified in each level based on PHQ-9 raw scores and based on DIF-corrected scores.

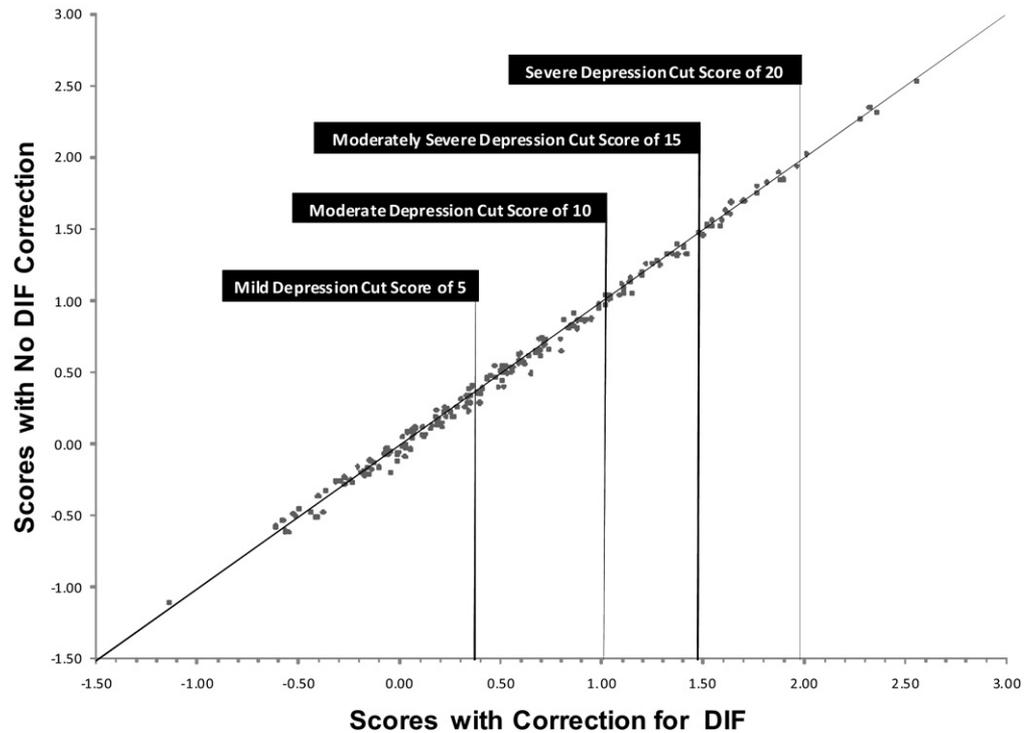


Fig 1. Comparison of IRT-calibrated scores (on logit scale) with and without correction for DIF. Logit values associated with summative score cutoff points of 5, 10, 15, and 20 are indicated on the x-axis.

The results indicate very small differences in the numbers of persons classified in each category, and the direction of differences depended on the range of depression being evaluated. For example, based on DIF-corrected scores, 2 more persons would be classified as having minimal depression and 2 more persons would be classified as having severe depression based on DIF-corrected scores.

To ensure that the disparity by sex of our primary care and TBI samples did not affect results, we combined the 2 samples and evaluated DIF by sex using our 3 criteria for meaningful DIF—change in pseudo- R^2 statistic less than .13, 5% change in beta, and 10% change in beta. None of the items of the PHQ-9 had meaningful DIF with respect to sex.

DISCUSSION

The results of this study do not support the notion that certain symptoms of DSM-IV major depression are invalid indicators among people with complicated mild to severe TBI. We found that in both the primary care sample and the TBI sample, PHQ-9 items loaded on a single depression factor. More importantly, when comparing PHQ-9 item functioning in the 2 samples, no PHQ-9 item demonstrated meaningful DIF attributable to TBI. Finally, the sensitivity analysis failed to show that the cumulative effects of negligible DIF resulted in

a systematic inflation of PHQ-9 total scores. Therefore, the findings do not support the hypothesis that cumulative DIF spuriously inflates diagnoses of MDD in people with TBI because of the influence of TBI-related symptoms that are not specific to MDD.

Taken together, the results suggest that clinicians should not eliminate or downplay cognitive and somatic symptoms of depression such as reduced energy, impaired concentration, and poor sleep in people with TBI. Rather, we recommend use of the “inclusive approach” to depression diagnosis. The inclusive approach means counting all DSM-IV symptoms of MDD toward the diagnosis regardless of whether the clinician judges the symptom to be caused by a medical condition (eg, TBI) or depression.²⁷ This approach is in contrast with the DSM-IV “etiologic” approach, in which the symptom is counted toward a diagnosis of major depression only if it is not fully accounted for by another medical condition. Among primary care patients as well as other medical populations, the inclusive approach is considered valid and perhaps even more reliable than other diagnostic strategies.²⁷⁻²⁹ Approaches that alter the diagnostic criteria depending on the medical condition are not recommended among primary care patients.²⁷

Prior research findings among people with TBI have been consistent with the inclusive approach. Jorge et al⁹ reported

Table 2: Comparison of Classifications of Depression Status Based on PHQ-9 Scores and DIF-Corrected Scores

Classification	Score Cutpoints		PHQ-9 Scores		Corrected for DIF	
	PHQ-9	Logit Scores	No.	Percentage	No.	Percentage
Minimal	0–4	<0.28	242	66.5	244	67.0
Mild	5–9	0.28–0.86	62	16.8	60	16.2
Moderate	10–14	0.87–1.36	31	8.5	33	9.1
Moderately severe	15–19	1.37–1.84	21	5.8	17	4.7
Severe	≥20	≥1.85	9	2.5	11	3.0

that patients with TBI who acknowledged having a depressed mood had higher rates of both physical and psychological symptoms of depression (as measured by the Present State Examination) compared with those without depressed mood. As a result, these investigators recommended the use of standard DSM-IV symptom criteria for diagnosing MDD in people with TBI.

In this study, we found no significant or clinically meaningful difference in the way persons with TBI respond to any of the PHQ-9 items compared with persons seen in primary care. The methodologic approach we used made it straightforward to demonstrate our results in the context of summative score criteria (eg, cutoff of 10). However, the implications of the findings extend to other PHQ-9–based algorithms for identifying presumptive diagnosis of depression. In previous research,¹⁷ we found a modified DSM-IV diagnostic approach to detecting MDD in persons with TBI was more accurate than use of a cut off score of 10. The current results reinforce the recommendation to use the inclusive approach in applying the DSM-IV diagnostic criteria.

To the extent that mental health professionals working in TBI have ignored or downplayed these transdiagnostic symptoms, major depression may have been underdiagnosed and undertreated. Unfortunately, research on depression treatment among people with TBI is consistent with this hypothesis. In a large prospective cohort of people with complicated mild to severe TBI, only 41% of people with evidence of a major depressive episode during the year after their injury received antidepressant medications, and 20% received any counseling during that year.¹ Overall, only 44% of those who were depressed received either antidepressants or counseling at any point during the first year after TBI. The degree to which underdiagnosis may contribute to the observed undertreatment is uncertain. Other factors such as poor access to treatment, cost of treatment, and stigmatization of mental health disorders and treatment may also contribute to undertreatment of depression among people with TBI. Nevertheless, systematic screening using an inclusive approach to diagnostic decisions may play a role in improving rates of depression treatment in this population. In previous research, we have shown that the PHQ-9 is a sensitive and specific screening tool for detecting DSM-IV MDD after TBI.¹⁷ While the field lacks definitive depression treatment efficacy studies,³⁰ we note that telephone-based psychosocial interventions³¹ and antidepressant treatment^{5,32} are promising approaches to treating depression-related suffering and disability after TBI.

Study Limitations

Several limitations of the study should be highlighted. The data on depression symptoms after TBI were gathered at a single urban level I trauma center in the Pacific Northwest. We urge caution in attempting to generalize these findings to regions with different sociodemographic characteristics. We used data gathered at 1 year after TBI. By this time, some of the TBI-related physical and cognitive symptoms would have abated. Greater DIF may be observed when depression is assessed closer to the time of injury. Studies are needed to explore whether DIF decreases as time since injury increases. The subjects in this study had sustained, complicated, mild to severe TBI. The results may not apply to people with uncomplicated mild TBI. Because we compared the TBI sample with primary care patients, we cannot rule out the possibility that people with TBI would demonstrate meaningful DIF compared with another comparison group. Nevertheless, we believe that it is an important step to demonstrate there is no meaningful DIF compared with primary care patients, a group within

which the validity of depression screening is less controversial.¹⁸

CONCLUSIONS

MDD is a prevalent, disabling, and undertreated condition associated with TBI. The use of valid and reliable depression screening tools is an important first step toward improved recognition and treatment. We have presented evidence that the PHQ-9 is a reliable and valid screener of MDD in people with complicated mild to severe TBI who are 1-year postinjury and that all the symptoms can be counted toward the diagnosis of MDD without special concern about overdiagnosis. We suggest that the PHQ-9 be used more widely in TBI care settings to identify those in need of further evaluation and potential treatment for depression.

APPENDIX 1: PHQ-9 ITEMS

-
1. Little interest or pleasure in doing things
 2. Feeling down, depressed, or hopeless
 3. Trouble falling or staying asleep, or sleeping too much
 4. Feeling tired or having little energy
 5. Poor appetite or overeating
 6. Feeling bad about yourself— or that you are a failure or have let yourself or your family down
 7. Trouble concentrating on things, such as reading the newspaper or watching television
 8. Moving or speaking so slowly that other people could have noticed. Or the opposite— being so fidgety or restless that you have been moving around a lot more than usual
 9. Thoughts that you would be better off dead, or of hurting yourself in some way
-

References

1. Bombardier CH, Fann JR, Temkin NR, Esselman PC, Barber J, Dikmen SS. Rates of major depressive disorder and clinical outcomes following traumatic brain injury. *JAMA* 2010;303:1938-45.
2. 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.
3. Rogers JM, Read CA. Psychiatric comorbidity following traumatic brain injury. *Brain Inj* 2007;21:1321-33.
4. Whitnall L, McMillan TM, Murray GD, Teasdale GM. Disability in young people and adults after head injury: 5-7 year follow up of a prospective cohort study. *J Neurol Neurosurg Psychiatry* 2006;77:640-5.
5. Fann JR, Uomoto JM, Katon WJ. Sertraline in the treatment of major depression following mild traumatic brain injury. *J Neuropsychiatry Clin Neurosci* 2000;12:226-32.
6. Rosenthal M, Christensen BK, Ross TP. Depression following traumatic brain injury. *Arch Phys Med Rehabil* 1998;79:90-103.
7. Kim E, Lauterbach EC, Reeve A, et al. Neuropsychiatric complications of traumatic brain injury: a critical review of the literature (a report by the ANPA Committee on Research). *J Neuropsychiatry Clin Neurosci* 2007;19:106-27.
8. Gass CS, Russell EW. MMPI profiles of closed head trauma patients: impact of neurologic complaints. *J Clin Psychol* 1991;47:253-60.
9. 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;181:91-9.
10. Moldover JE, Goldberg KB, Prout MF. Depression after traumatic brain injury: a review of evidence for clinical heterogeneity. *Neuropsychol Rev* 2004;14:143-54.

11. Babin PR. Diagnosing depression in persons with brain injuries: a look at theories, the DSM-IV and depression measures. *Brain Inj* 2003;17:889-900.
 12. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997;6:595-600.
 13. Hambleton RK. *Fundamentals of item response theory*. Newbury Park: Sage Publications, Inc; 1991.
 14. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-13.
 15. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002;32:509-21.
 16. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry* 2010;32:345-59.
 17. Fann JR, Bombardier CH, Dikmen S, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil* 2005;20:501-11.
 18. 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;282:1737-44.
 19. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45(Suppl 1):S22-31.
 20. Samejima F. Estimation of latent ability using a response pattern of graded scores (psychometric monograph no. 17). Richmond: Psychometric Society; 1969.
 21. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Med Care* 2006;44(Suppl 3):S115-23.
 22. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates; 1988.
 23. Zumbo BD. A handbook on the theory and methods of differential item functioning (DIF): logistic regression modeling as a unitary framework for binary and Likert-type(ordinal) item scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.
 24. Crane PK, van Belle G, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. *Stat Med* 2004; 23:241-56.
 25. Crane PK, Gibbons LE, Ocepek-Welikson K, et al. A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Qual Life Res* 2007;16(Suppl 1):69-84.
 26. Faul M, Xu L, Wald MM, Coronado VG. *Traumatic brain injury in the United States: emergency department visits, hospitalizations and deaths, 2002–2006*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010.
 27. Williams JW, Jr, Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA* 2002;287:1160-70.
 28. Koenig HG, George LK, Peterson BL, Pieper CF. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. *Am J Psychiatry* 1997;154:1376-83.
 29. Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. *Psychol Med* 2006;36:27-36.
 30. Fann JR, Hart T, Schomer KG. Treatment for depression after traumatic brain injury: a systematic review. *J Neurotrauma* 2009; 26:2383-402.
 31. Bombardier CH, Bell KR, Temkin NR, Fann JR, Hoffman J, Dikmen S. The efficacy of a scheduled telephone intervention for ameliorating depressive symptoms during the first year after traumatic brain injury. *J Head Trauma Rehabil* 2009;24:230-8.
 32. Novack TA, Banos JH, Brunner R, Renfro S, Meythaler JM. Impact of early administration of sertraline on depressive symptoms in the first year after traumatic brain injury. *J Neurotrauma* 2009;26:1921-8.
- Suppliers**
- a. Muthén & Muthén, 3463 Stoner Ave, Los Angeles, CA 90066.
 - b. The Comprehensive R Archive Network, Available at: <http://cran.r-project.org>.