



Published in final edited form as:

*Rehabil Psychol.* 2009 November ; 54(4): 440. doi:10.1037/a0017402.

## Comparison of the Patient Health Questionnaire and the Older Adult Health and Mood Questionnaire for self-reported depressive symptoms after spinal cord injury

James S. Krause, Lee L. Saunders, Karla S. Reed, Jennifer Coker, Yusheng Zhai, and Emily Johnson

Medical University of South Carolina

### Abstract

**Objective**—To directly compare estimates of potential depressive disorders and clinically significant depressive symptoms using the Patient Health Questionnaire (PHQ-9) and Older Adult Health and Mood Questionnaire (OAHMQ) among participants with spinal cord injury (SCI).

**Research Design**—727 participants from a hospital in the Southeastern United States were administered the PHQ-9 and OAHMQ during a follow-up survey. We compared the rates of depressive disorders using cutoff scores and diagnostic criteria for each instrument. No independent psychiatric diagnostic interviews were conducted.

**Results**—The PHQ-9 and OAHMQ were significantly correlated ( $r=.78$ ) and both were correlated with satisfaction with life ( $r=-.48, -.54$ ). Using recommended diagnostic scoring procedures, 10.7% of participants met the diagnostic criteria for major depressive disorder with the PHQ-9, 9.3% met the criteria for major depression based on PHQ-9  $\geq 10$ , and 19.7% based on PHQ-9  $\geq 15$ . Using the OAHMQ, 19.7% reported probable major depression and 44.5% clinically significant symptomatology.

**Conclusions**—The measures were highly correlated overall. However, the estimated prevalence of depressive disorders varied substantially between the two instruments. These estimates were comparable to those previously reported for each instrument (i.e., higher rates with the OAHMQ). Therefore, differing estimates of depressive disorders reported in the literature using these instruments were largely attributable to the instruments themselves.

### Keywords

spinal cord injury; depression; psychometrics

---

A substantial number of studies have been conducted in order to identify the association of neurologic injury, including spinal cord injury (SCI), with depressive disorders. These studies have produced wide variations in rates of depressive disorders. For example, clinical practice guidelines published in 1998 estimated that as many as 25% of men and 47% of women with

---

Correspondence concerning this article should be addressed to James S. Krause, College of Health Professions, 77 President St, Suite 117, Charleston, SC 29425. krause@musc.edu.

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/pubs/journals/rep](http://www.apa.org/pubs/journals/rep).

SCI experienced some form of depressive disorder (Consortium for Spinal Cord Medicine, 1998). Data reported in two reviews of the literature indicated rates of major depression ranging from 11% to over 30% (Elliott & Frank, 1996; Frank, Elliott, Corcoran, & Wonderlich, 1987).

One of the likely reasons for the discrepancies in rates of depressive disorders relates to variations in the specific measures used between studies and their scoring criteria. Yet, there are few studies that anchor estimates from one instrument with those of another due to the utilization of a single instrument in most studies. Direct comparisons of instruments within the same study would essentially enhance our ability to interpret differential estimates of depressive disorders.

A number of different measures of depressive disorders have been used in recent studies with SCI, including the Beck Depression Inventory (Oh, Shin, Paik, Yoo, & Ku, 2006; Pollard & Kennedy, 2007), the Center for Epidemiological Studies Depression Scale (Anton, Miller, & Townson, 2008; Miller, Anton, & Townson, 2008), Depression Anxiety Stress Scales-21 (Migliorini, Tonge, & Taleporos, 2008; Mitchell, Burns, & Dorstyn, 2008), Hospital Anxiety and Depression Scale (Woolrich, Kennedy, & Tasiemski, 2006), Older Adult Health and Mood Questionnaire (OAHMQ; Kemp & Adams, 1995), Patient Health Questionnaire-9 (PHQ-9; Graves & Bombardier, 2008; Kalpakjian & Albright, 2006; Krause, Bombardier, & Carter, 2008; Richardson & Richards, 2008), the Questionnaire for Depression (Gioia, et al., 2006), the Spinal Cord Lesion Coping Strategies Questionnaire (Elfstrom, Ryden, Kreuter, Persson, & Sullivan, 2002; Migliorini, Elfstrom, & Tonge, 2008), and the Spinal Cord Lesion Emotional Well-Being Questionnaire (Elfstrom, et al., 2002; Migliorini, Elfstrom, et al., 2008).

The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) has become the focus of many studies with SCI, at least partially due to its inclusion in the SCI Model Systems (NSCISC, 2005). It is directly anchored by the Diagnostic and Statistical Manual of Mental Disorders-IV criteria (DSM-IV; APA, 1994), as each of the items reflects a specific DSM-IV depressive symptom, including several somatic symptoms. Despite it being the most widely used measure in recent studies, it is important to base decisions regarding the measure of choice on empirical data, rather than frequency of utilization or the availability of data from existing sources, such as the SCI Model Systems (NSCISC, 2005).

The PHQ-9 utilizes two alternative scoring systems for diagnostic procedures for major depression. One scoring system involves comparing responses directly to diagnostic criteria from the DSM-IV (APA, 1994), requiring endorsement of one of two cardinal symptoms (“having little interest or pleasure in doing things” or “feeling down, depressed or hopeless”), and five out of nine items must be reported on half or more days. The lone exception is the item on self-harm (suicidal ideation) which is included if any days are endorsed. The second method is to use a cutoff of 10 or 15, indicating probable major depressive disorder (MDD; Kroenke, et al., 2001)). This second method has been found to be highly valid when compared to a structured clinical interview. Although both cutoffs appear acceptable, 10-14 was considered a “gray zone.”

Studies using the PHQ-9 have produced substantially lower rates of major depression among SCI samples relative to the rates reported with other measures of depression. In a study on SCI Model Systems participants, Bombardier et al. (2004) found that 22% scored at or above the cutoff for MDD, using a criterion of 10 or higher (cutoff scores can be either 10 or 15). Just over 11% of the sample met the criteria for probable major depression. In another study, investigators found a prevalence rate of probable MDD of 7.9% for women and 9.9% for men (Kalpakjian & Albright, 2006).

Recent studies have suggested a 2-factor structure for the PHQ-9 with a distinctive somatic factor (Krause, et al., 2008; Richardson & Richards, 2008), although some studies have suggested that a single factor best fits the data (Kalpakjian et al., in press). Graves and Bombardier (2008) recently presented data from confirmatory factor analysis that a 1-factor solution produced a root mean square error of approximation (RMSEA) of 0.091 (CI=0.086, 0.097), which the authors cite as a “fair” fit using the liberal rule-of-thumb of 0.10 (lower RMSEA indicates a better fit). This exceeds the criterion of 0.080 or lower for a good fit (Browne & Cudeck, 1993) and is substantially higher than that of a 2-factor model (RMSEA = 0.073) in a study (Krause, et al., 2008) that directly compared 1-factor and 2-factor models among participants during inpatient hospitalization (Graves and Bombardier only tested a single model).

Although not used as frequently as the PHQ-9 in recent SCI studies, the OAHMQ (B.J. Kemp & Adams, 1995) is another viable measure, as it is specifically for use among older participants and those with disabling conditions. The OAHMQ is a 22-item true/false measure designed to include few vegetative items. Cutoff scores have been validated with clinical diagnoses. Separate cutoff scores are included for clinically significant symptomatology (CSS) (6-10) and probable major depression (PMD; 11-22). The OAHMQ uses a three part division of non-depressed, CSS, and PMD and is found to have a sensitivity of .80 and specificity of .87 for major depression.

Several studies have been conducted using the OAHMQ with SCI populations. As a whole, these studies have produced relatively high rates of depressive disorders. For instance, in one study, 48% of participants with SCI were found to have CSS, with half of those showing signs of PMD (Krause, Kemp, & Coker, 2000). Similarly, (Kemp & Krause, 1999) found that 22% of the post-polio group, 41% of the SCI group, and 15% of a group without a disabling condition had at least significant depressive symptomatology (scores of 6 or higher). In a study of participants with SCI from three different ethnic groups (African Americans, Latinos, and Caucasians), 42% of the participants reported levels of CSS, with 18% reporting PMD (B. Kemp, Krause, & Adkins, 1999).

A factor analysis of the OAHMQ identified four factors (Krause, et al., 2000), the first three of which were used for scale development. Three scales were developed, which were labeled as follows: (1) Evaluative; (2) Affective; and (3) Behavioral. Each of the three factors represented a particular aspect of depressive symptoms. The first factor reflected a general negative evaluation of life, including a gloomy outlook on life, a loss of interest or pleasure, and a sense of hopelessness about the future. In contrast, the second factor reflected the affective or mood component of the OAHMQ, with items of sadness and tearfulness. The third factor mostly reflected a behavioral component with a change toward fewer activities. The factors were similar to those identified in an earlier SCI study by Kemp and colleagues (B. Kemp, et al., 1999).

## Purpose

The purpose of this study is to directly compare rates of potential depressive disorders or clinically significant depressive symptoms using two measures -- the PHQ-9 and the OAHMQ among participants with SCI. We chose these two measures because of their complementary strengths and weaknesses, utilization in several SCI studies, and their differing estimates of depressive disorders reported in the literature. The PHQ-9 has the advantage of being developed for the general population, items based on DSM-IV criteria (APA, 1994), and use in the SCI Model Systems data collection. In contrast, the OAHMQ has the advantage of being specifically developed for individuals with health concerns and disabilities and inclusion of a limited number of vegetative items. Identifying the varying estimated rates of depressive

disorders between the two instruments in the same study will allow investigators to directly compare alternative cutoff scores between the two instruments.

### Research Questions

1. To what extent do the two measures correspond in terms of depressive diagnoses using traditional cutoff scores (i.e., those proposed by the test developers)?
2. Assuming that there are discrepancies in rates of diagnoses between the two measures, what adjustments would be required in order for the two measures to produce similar estimates?
3. To what extent are depressive symptomatology scores from each measure correlated with satisfaction with life?

### Methods

#### Participants

After receiving approval from the Institutional Review Board, participants were identified from records of a specialty hospital in the Southeastern United States. All participants were adults with traumatic SCI of at least one year duration. A total of 1,385 participants were enrolled in the original study in 1997-1998. Participants were then contacted in 2007-2008 to participate in a follow-up survey. At that time, 306 were deceased, 34 were lost (could not be located), and 5 were eliminated. Responses were received by 727 participants, yielding an adjusted response rate of 69.5% percent.

The majority of participants were Caucasian (75.8%) and male (70.2%). Motor vehicle crashes were the primary etiology (51.5%), followed by falls/flying objects (13%), sporting injuries (11.2%), and acts of violence (9.2%). Over half of the participants (53.3%) had cervical injuries. Neurologic completeness of injury was broken down into four groups similar to, but not identical, to the ASIA grades (Maynard, et al., 1997). Just over 40% (41.8) reported no movement or sensation below the level of injury, 20.6% reported sensation only, 8.7% reported non-functional motor recovery, and 21.2% were ambulatory. The average age at the time of the study was 47.9 and average number of years since injury was 18.2. The average number of years of education was 13.8.

#### Procedures

Participants were initially contacted through a preliminary letter describing the study and alerting them that materials would be forthcoming. Updated addresses were requested from the United States Post Office for those who had moved recently. Four to six weeks later, an initial packet of materials was mailed to participants followed by a second set of materials for all non-respondents. We also called non-respondents to elicit participation and sent out additional materials to those who had misplaced or discarded materials but consented to participate by phone and requested an additional set of materials. Participants were offered \$50 remuneration for their participation in the study.

#### Measures

The PHQ-9 was developed as a scaled down version of the PRIME-MD, an instrument used to diagnose common mental disorders (Kroenke, et al., 2001). Specifically, the PHQ-9 was used to diagnose depressive disorders (MDD or Other Depressive Disorder) by using nine items, each identifying symptoms commonly associated with a depression diagnosis (e.g., having little interest or pleasure in doing things, feeling down, depressed, or hopeless). The participant was asked to identify how frequently each symptom has been a problem over the

past two weeks using four categories: a) not at all; b) several days; c) more than half of the days; and d) nearly every day. Severity of depression was categorized two different ways. First, if the participant answered “more than half the days” on five of the nine items with one of those five items either “having little interest or pleasure in doing things” or “feeling down, depressed or hopeless,” then the person was categorized as MDD. Several days or more was used as the criteria for the item “Thoughts of being better off dead or of hurting yourself in some way.” Second, the PHQ-9 was assessed as a continuous variable, with scores across all nine variables being summed to yield a continuous measure of depression severity. Two cut points were used for a depressive disorder (10 and 15), consistent with that reported in earlier research by Bombardier et al. (2004). They found preliminary evidence that this questionnaire may be a useful screen for depression in people with SCI. The measure showed high internal consistency and strong item-total correlations. Also, Kroenke et al. (2001) found .89 internal consistency and .84 test-retest reliability.

The OAHMQ (B.J. Kemp & Adams, 1995) was designed to evaluate depression in older adults and among people with physical disabilities as a 22-item questionnaire with few items reflecting physical or vegetative symptomatology. These types of items often invalidate commonly used measures of depression and other clinical syndromes with people with SCI. Previous research indicates that these types of symptoms may not truly reflect depression rather the physiologic changes brought about by the injury itself (Taylor, 1967). The items of the OAHMQ were developed as true/false statements such as “My daily life is interesting” or “I still have regrets about the past that I think about often.” Scores of 6-10 were considered CSS, whereas scores of 11 and higher were considered PMD.

Recent work by the instrument's author indicated that the OAHMQ may be sensitive to race-ethnicity, as Hispanics scored lower on the OAHMQ than both Caucasians and African-Americans (B. Kemp, et al., 1999). Test-retest correlations were found to be .87 ( $p < .001$ ) and alphas were .93 ( $p < .001$ ). In terms of validity, sensitivity was found to be .93 and specificity was .87. The OAHMQ was also compared to two other depression scales with established validity and reliability, the Geriatric Depression Scale (Yesavage, et al., 1983) and the depression scale from the Symptom Checklist-90-revised (Derogatis, 1992). The OAHMQ was correlated .70 with each measure.

The Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) is a 5 item measure of life satisfaction. Participants were asked to identify their agreement with five statements using the seven categories: a) strongly disagree, b) disagree, c) slightly disagree, d) neutral, e) slightly agree, f) agree, and g) strongly agree. Scores ranged from 5-35, with higher scores indicating a higher level of satisfaction with life. Multiple studies have reported high internal reliability for the SWLS with alpha coefficients ranging from .79 to .87 (Blais, Vallerand, Pelletier, & Briere, 1989; Diener, et al., 1985; Pavot, Diener, Colvin, & Sandvik, 1991; Yardley & Rice, 1991). Convergent validity was noted by correlations with other measures of life satisfaction and subjective well-being ranging from .35 to .60 (Diener, et al., 1985; Frisch, Cornell, Villanueva, & Retzlaff, 1992; Larsen, Diener, & Emmons, 1985).

## Analyses

Descriptive statistics were calculated for all biographic and injury-related variables. Continuous variables were expressed as *mean (standard deviation)*, while categorical variables were expressed as *N* (proportion). Spearman's rank correlation coefficients were calculated between two depressive disorders scores (PHQ-9 and OAHMQ), their cut points, and the SWLS. Cronbach's Alpha coefficients were calculated in order to check the internal consistency reliability within each depressive disorder scale.

Three logistic regression models were run using the three binary cut points for the PHQ-9 as the dependent variable and the OAHMQ as the independent variable. The PHQ-9 was selected as the dependent variable because it is anchored against DSM-IV criteria (APA, 1994). From the logistic models, Receiver Operating Characteristic (ROC) plots and area under the curve (AUC) were used to select possible optimal cut points for the OAHMQ using the three PHQ-9 cut points (MDD, PHQ-9  $\geq$  10, PHQ-9  $\geq$  15) as the standard (Hosmer & Lemeshow, 2000). Sensitivity and specificity combined with the Youden's J statistic were calculated to decide the best cut point for OAHMQ when compared to the two PHQ-9 cut points.

## Results

### Summary of PHQ-9, OAHMQ

The average score on the PHQ-9 was 5.57 ( $SD=5.74$ ). Table 1 presents a breakdown of PHQ-9 scores using the same categories as Bombardier et al. (2004), and also using the cut points from the continuous variable (both 10 and 15). Just over 10% of the participants met the diagnostic criteria for MDD. Additionally, 19.7% met the PHQ-9  $\geq$  10 diagnostic criterion. When using a cutoff score of 15, only 9.3% were above the cutoff for moderately severe to severe depression. The alpha coefficient for the full scale was .89.

The average score on the OAHMQ was 6.0 ( $SD=5.0$ ). When using the recommended cutoff of 6 or higher for CSS, just less than half of the participant sample (44.5%) met the criterion for CSS (see Table 1). This figure decreased substantially (19.7%) when using scores of 11 or higher for PMD. The alpha coefficient for the full scale was .87.

Both the PHQ-9 and the OAHMQ were broken down by race in Table 1. The PHQ-9 did not differ by race ( $p > .05$ ); however, the OAHMQ was significantly different between the races,  $\chi^2(2, N = 682) = 6.2, p = .044$ . Participants who were white were more likely to report no depressive symptoms than non-whites.

### Correspondence between Scoring Methods

Table 2 presents the Spearman rank correlations among the PHQ-9 and OAHMQ scoring methods and SWLS. All correlations were statistically significant at  $p < .001$ . The two continuous measures of depression were significantly positively correlated  $r(655) = .78$ . The OAHMQ was slightly more highly correlated with the SLWS,  $r(671) = -.54$  than was the PHQ-9,  $r(678) = -.48$ . The alpha coefficient for the SWLS was .92.

When comparing the mean of the SWLS as a function of diagnostic cutoff points (Table 3), the mean SLWS was lower among those with MDD on the PHQ-9 (11.87 compared with 21.67). SWLS scores were also significantly different as a function of cutoff scores of both 10 (13.22, 22.4) and 15 point cutoffs (10.94, 21.62), as well as for the OAHMQ  $\geq$  11 (12.46 compared with 22.53).

The relative extent of agreement in diagnoses using the various diagnostic cutoffs between the two measures ranged substantially depending on which cutoff scores were used. The index of agreement indicates the correspondence between diagnoses using two alternative cutoffs (i.e., the extent to which the same individuals were above or below each cutoff). The OAHMQ cutoff for PMD ( $\geq$  11) was more correlated with MDD (.56) and PHQ-9  $\geq$  10 (.69) than with PHQ-9  $\geq$  15 (.48).

Table 4 presents data comparing classifications based on the different scoring procedures of the PHQ-9 and OAHMQ. Although no true diagnostic gold standard was available to us, we have treated the scoring of the PHQ-9 as such a standard, based on its widespread usage and ties with DSM-IV criteria (APA, 1994). This permits us to describe our results in familiar terms

of sensitivity and specificity in describing the correspondence between scoring procedures and in presenting the ROC curve analyses described below. The OAHMQ  $\geq 11$  captures 90% of those with MDD, 86% of those with PHQ-9  $\geq 15$ , and 75% of those with PHQ-9  $\geq 10$  (Table 4). In addition, the OAHMQ  $\geq 11$  has fewer false positives with the PHQ-9  $\geq 10$  (6%) and the MDD (11%) than the PHQ-9  $\geq 15$  (13%). Kappa statistics showed the highest agreement between OAHMQ and PHQ-9 cut points was OAHMQ  $\geq 11$  and PHQ-9  $\geq 10$ .

A series of ROC curves were also generated to explore the best cut point for the OAHMQ when examined against the PHQ-9 (treating this as the gold standard). Figure 1 demonstrates the ROC curve for the OAHMQ compared with MDD. With a sensitivity of 86.8% and a specificity of 92.7%, a value of OAHMQ=12 was shown to be the best cut point for MDD.

Figure 2 demonstrates the ROC curve for the OAHMQ compared with PHQ-9  $\geq 15$ . With a sensitivity of 91.4% and a specificity of 85.1%, a value of OAHMQ=10 was shown to be the best cut point for PHQ-9  $\geq 15$ .

Lastly, figure 3 demonstrates the ROC curve for the OAHMQ compared with PHQ-9  $\geq 10$ . A value of 85.9% and specificity of 89.6% were shown by the best cut point for PHQ-9  $\geq 10$ , OAHMQ=9.

## Discussion

The results of this study help to clarify the comparability of the PHQ-9 and OAHMQ when used with SCI. The two measures were highly correlated overall. Similarly, they were both modestly and significantly correlated with life satisfaction, as would be expected, but these correlations (-.48 and -.54 respectively) were less than the two measures were correlated with each other (+0.78).

This study also clarified discrepancies in estimates of depressive diagnoses cited in previous research of SCI using these measures, as we found widely differing estimates of depressive diagnoses ranging from a low of 9.3% (using cutoff of 15 with the PHQ-9) to a high of 44.5% of CSS using the OAHMQ *among the same participants*. When restricting comparisons to MDD (from the PHQ-9) and PMD (OAHMQ  $\geq 11$ ), relatively comparable diagnostic categories, the discrepancies were not as profound yet substantial (10.7% compared with 19.7%).

The rates of depressive disorders in the current study using the PHQ-9 were similar to those reported previously by Bombardier et al. (2004) who used the PHQ-9 with data from the SCI Model Systems. They reported 11.4% of their participants met the criteria for MDD (compared with 10.7% in the current study) and 22% were at the cutoff score of 10 (compared with 19.7% in the current study). Therefore, despite substantial differences in participant identification between our study and the Bombardier study, the diagnostic rates using the PHQ-9 were highly similar.

Similarly, for the OAHMQ, our rates of CSS (44.5%) and PMD (19.7%) were similar to estimates from other studies reported in the literature. For instance, CSS ranged from 41-48% in the aforementioned studies, whereas the rate of PMD was 18% (B. Kemp, et al., 1999; B J Kemp & Krause, 1999; Krause, et al., 2000).

In summary, although our estimates of depressive disorders *between* the PHQ-9 and the OAHMQ were substantial, they were consistent with those from other studies reported in the literature that used only one or the other instrument. In other words, the differences we observed in the portion of depressive disorders are likely a function of the instruments themselves when used with SCI.

It is important to note that, although the diagnostic cutoffs produced very different estimates of depressive disorders, when using alternative cutoffs, the diagnostic rates between instruments were brought in line with each other. For instance, a cutoff of 12 on the OAHMQ resulted in a sensitivity of 86.8% and a specificity of 92.7% with MDD. Similarly, a cutoff point of 10 on the OAHMQ maximized the correspondence with the PHQ-9 cutoff of 15, and a score of 9 on the OAHMQ corresponded with a cutoff score of 10 on the PHQ-9.

### Implications

These results help guide clinicians as to the limitations of the two self-report measures, rather than providing an absolute statement as to utility. Because the PHQ-9 was designed to collect the diagnostic criteria from the DSM-IV (APA, 1994), it may be the measure of choice when the screening is related to diagnosis. However, given its reliance on somatic content and the profound physiologic changes with SCI, there will be populations and circumstances where it is preferable to use the OAHMQ.

Although it is tempting to make absolute statements regarding which cutoff scores to use, it is best that this is determined by the goal of the assessment. If the goal is to identify the likelihood of a depressive disorder, then it is probably best to use the MDD of the PHQ-9 or the PMD of the OAHMQ. However, if the goal is to identify individuals who are at risk for general adverse outcomes due to the presence of mild symptoms or those who may be at risk for developing MDD in the future, then it would be more appropriate to use the much more liberal CSS cutoff of the OAHMQ. Regardless of intention or instrument, if the ultimate goal is to diagnose depressive disorders, then a clinical interview is required (the PHQ-9 and OAHMQ are screening and not diagnostic measures).

### Limitations

There are several limitations in this study. First, there were no independent measures of depressive disorders, such as diagnostic interviews. This information would additionally help us to compare both measures to an independent diagnostic criterion, rather than simply to each other. It is noteworthy that one of the primary reasons for utilizing self-report measures is that diagnostic interviews are labor intensive and not feasible for epidemiological studies. Second, we do not have data regarding types of treatments that individuals may have received or whether they were currently being treated. It would be interesting to know whether individuals were receiving psychotherapy or were being treated with medication. Third, we used only two of a large number of possible measures. Therefore, this study does not clarify variations in rates of depressive diagnoses using other measures. Similarly, our results apply only to a single population -- SCI. Lastly, the data are cross-sectional, so we do not know how scores on one measure predict future scores or outcomes. Differences in predictive validity would be very important in assessing the practical utility of a measure for research or practice.

### Future Research

Additional research is needed directly comparing multiple measures of depressive symptoms and diagnoses or multiple populations within the same study. Another important direction for future research is to identify not only the current value of these screening measures but also their predictive value of future depressive episodes and other associated outcomes. It may be that one measure better predicts future outcomes than another, which would be an important discovery for applications where the goal is to predict future outcomes. Lastly, more research is needed on the quality of different measures within treatment studies. Although randomized clinical trials generally utilize detailed assessments, it may be valuable to also identify how relatively short screening measures are able to capture changes in outcomes. These are several directions for future research, all of which have the potential to improve our understanding of the utility of screening measures for depression with special populations.



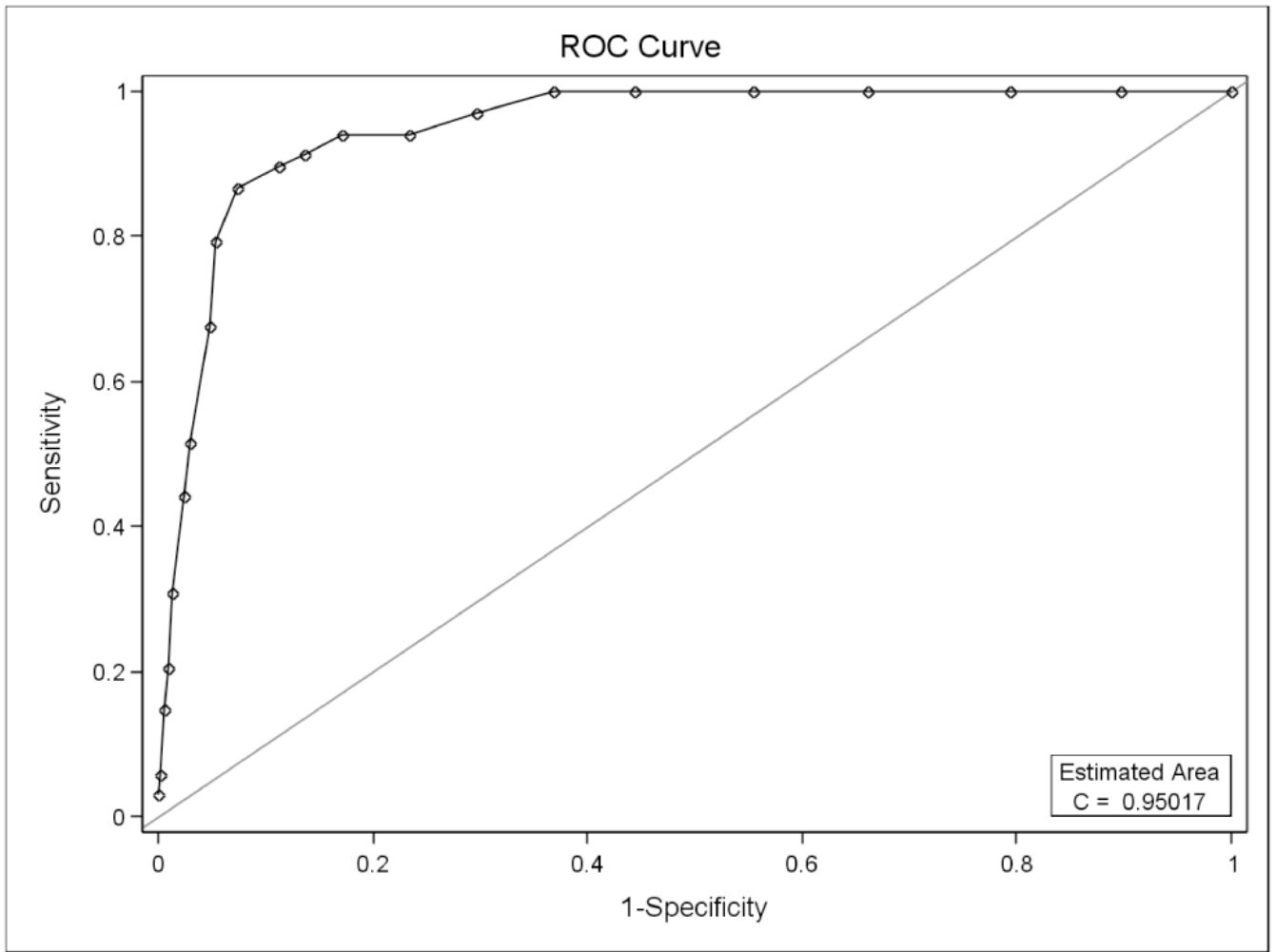
## Acknowledgments

This research was supported by a field initiated grant from the National Institute for Disability and Rehabilitation Research (H133G050165), the Model Spinal Cord Injury Systems Grant (H133N000005, H133N060009), and the National Institute of Health (1R01 NS 48117). The opinions here are those of the grantee and do not necessarily reflect those of the funding agencies.

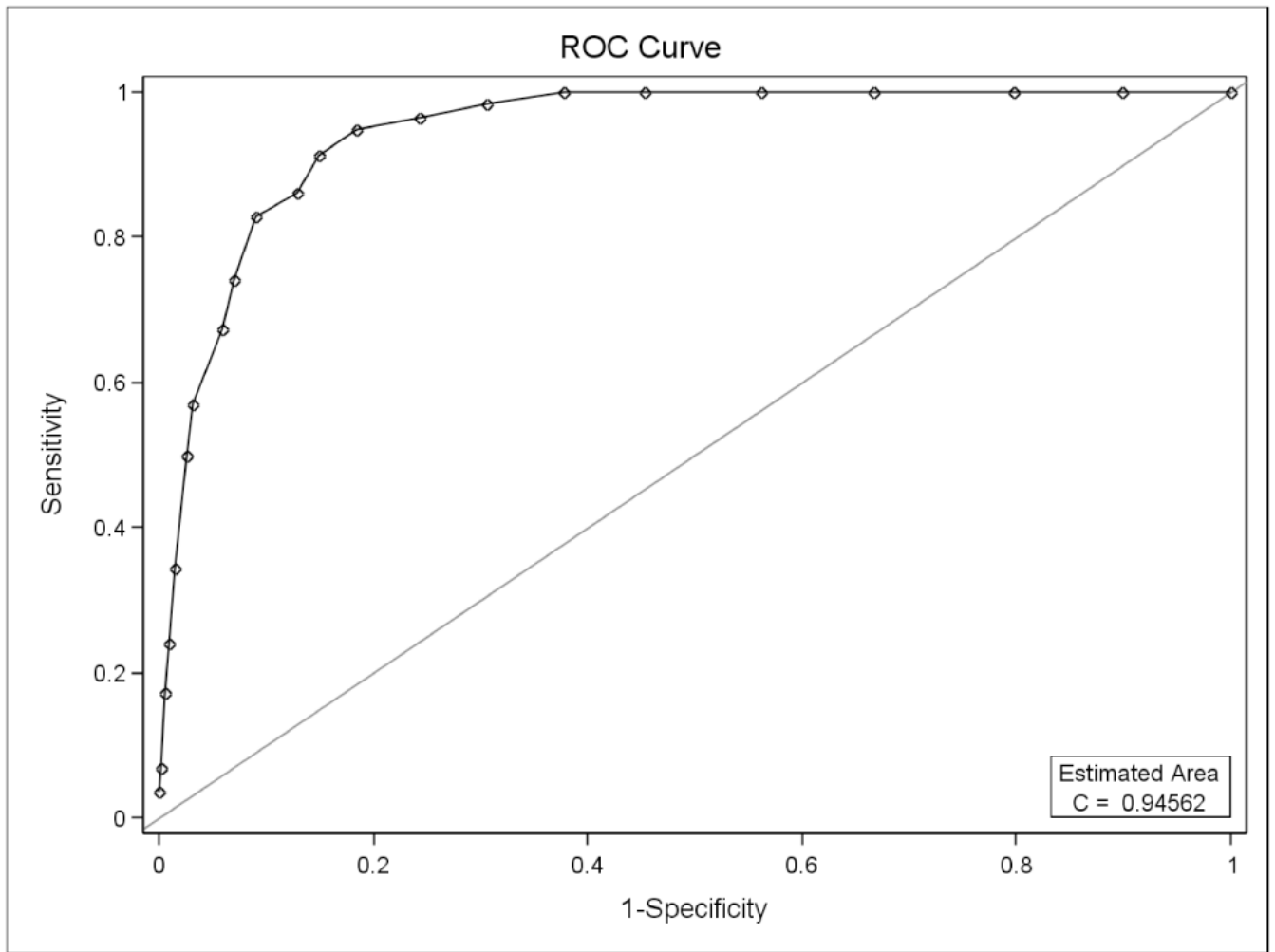
## References

- Anton HA, Miller WC, Townson AF. Measuring fatigue in persons with spinal cord injury. *Archives of Physical Medicine and Rehabilitation* 2008;89:538–542. [PubMed: 18295634]
- APA. *Diagnostic and statistical manual of mental disorders*. 4th. Washington, DC: American Psychiatric Association; 1994.
- Blais MR, Vallerand RJ, Pelletier LG, Briere NM. L'Echelle de satisfaction de vie: Validation Canadienne-Francaise du "Satisfaction With Life Scale" [French-Canadian Validation of the Satisfaction With Life Scale]. *Canadian Journal of Behavioral Science* 1989;21:210–223.
- Bombardier CH, Richards JS, Krause JS, Tulsy D, Tate DG. Symptoms of major depression in people with spinal cord injury: implications for screening. *Archives of Physical Medicine and Rehabilitation* 2004;85(11):1749–1756. [PubMed: 15520969]
- Browne, M.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, K.; Long, S., editors. *Testing Structural Equation Models*. Beverly Hills, CA: Sage; 1993. p. 136-162.
- Consortium for Spinal Cord Medicine. Washington, DC: Paralyzed Veterans of America; 1998.
- Derogatis, LR. *The Symptom Checklist-90-revised*. Minneapolis, MN: NCS Assessments; 1992.
- Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *Journal of Personality Assessment* 1985;49:71–75. [PubMed: 16367493]
- Elfstrom ML, Ryden A, Kreuter M, Persson LO, Sullivan M. Linkages between coping and psychological outcome in the spinal cord lesioned: development of SCL-related measures. *Spinal Cord* 2002;40:23–29. [PubMed: 11821966]
- Elliott TR, Frank RG. Depression following spinal cord injury. *Archives of Physical Medicine & Rehabilitation* 1996;77(8):816–823. [PubMed: 8702378]
- Frank RG, Elliott TR, Corcoran JR, Wonderlich SA. Depression after spinal cord injury: Is it necessary? *Clinical Psychology Reviews* 1987;7:611–630.
- Frisch MB, Cornell J, Villanueva M, Retzlaff P. Clinical validation of the Quality of Life Inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment* 1992;4:92–101.
- Gioia MC, Cerasa A, Di Lucente L, Brunelli S, Castellano V, Trallesi M. Psychological impact of sports activity in spinal cord injury patients. *Scandinavian Journal of Medicine & Science in Sports* 2006;16:412–416. [PubMed: 17121643]
- Graves DE, Bombardier CH. Improving the efficiency of screening for major depression in people with spinal cord injury. *Journal of Spinal Cord Medicine* 2008;31:177–184. [PubMed: 18581665]
- Hosmer, DW.; Lemeshow, S. *Applied logistic regression*. 2nd. New York, New York: John Wiley and Sons; 2000.
- Kalpakjian CZ, Albright KJ. An examination of depression through the lens of spinal cord injury: Comparative prevalence rates and severity in women and men. *Women's Health Issues* 2006;16:380–388. [PubMed: 17188221]
- Kemp B, Krause J, Adkins R. Depression among African Americans, Latinos, and Caucasians with spinal cord injury: A exploratory study. *Rehabilitation Psychology* 1999;44(3):235–247.
- Kemp BJ, Adams BM. The Older Adult Health and Mood Questionnaire: A measure of geriatric depressive disorder. *Journal of Geriatric Psychiatry and Neurology* 1995;8(3):162–167. [PubMed: 7576040]
- Kemp BJ, Krause JS. Depression and life satisfaction among people aging with post-polio and spinal cord injury. *Disability and Rehabilitation* 1999;21:241–249. [PubMed: 10381236]

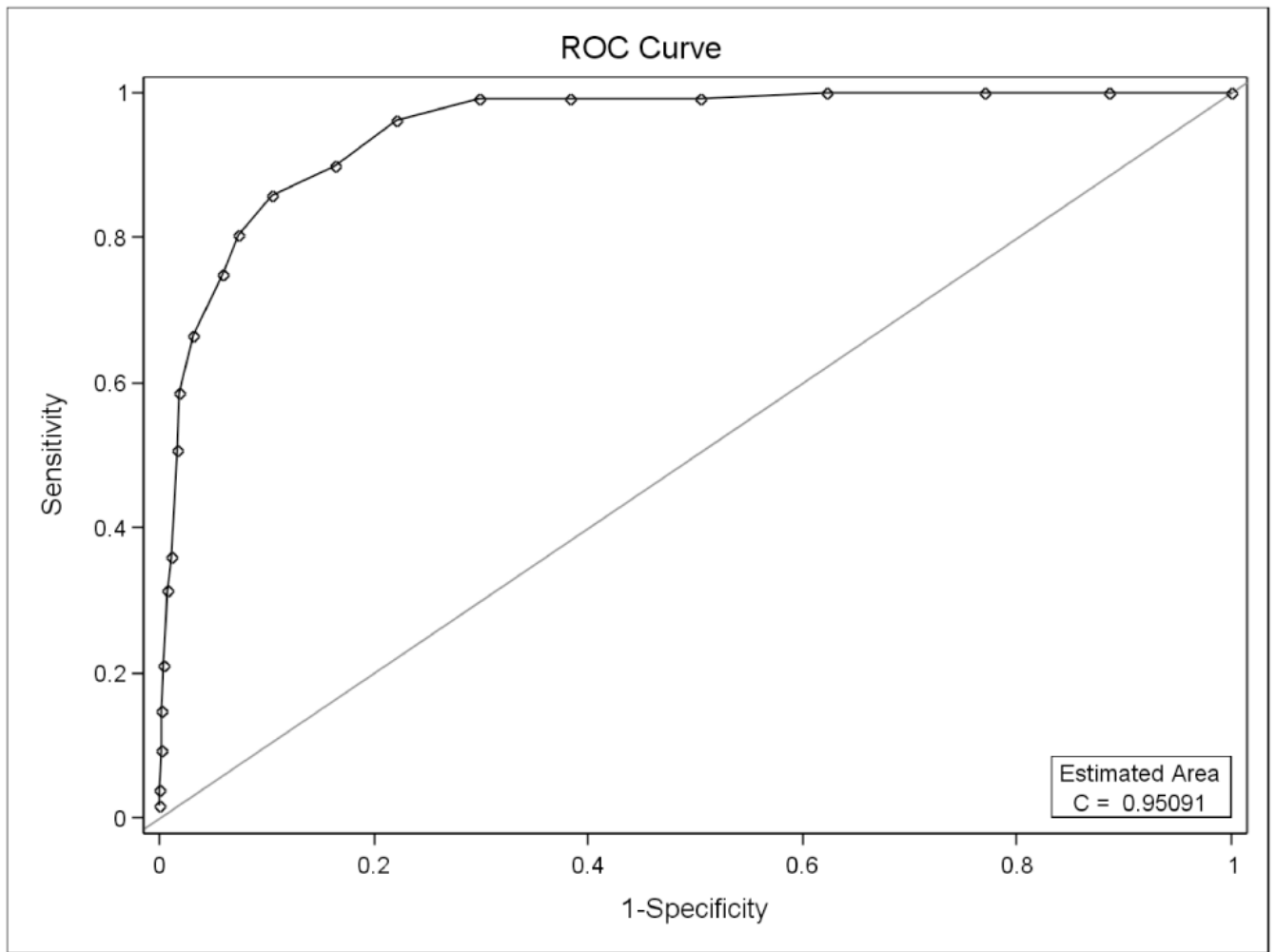
- Krause JS, Bombardier CH, Carter RE. Assessment of depressive symptoms during inpatient rehabilitation for spinal cord injury: Is there an underlying somatic factor when using the PHQ? *Rehabilitation Psychology* 2008;53(4):513–520.
- Krause JS, Kemp B, Coker JL. Depression after spinal cord injury: relation to gender, ethnicity, aging, and socioeconomic indicators. *Archives of Physical Medicine and Rehabilitation* 2000;81(8):1099–1109. [PubMed: 10943762]
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 2001;16(9):606–613. [PubMed: 11556941]
- Larsen RJ, Diener E, Emmons RA. An evaluation of subjective well-being measures. *Social Indicators Research* 1985;17:1–18.
- Maynard FM, Bracken MB, Creasey G, Ditunno JF, Donovan WH, Ducker TB, et al. International standards for neurological and functional classification of spinal cord injury. *American Spinal Injury Association. Spinal Cord* 1997;35:266–274. [PubMed: 9160449]
- Migliorini CE, Elfstrom ML, Tonge BJ. Translation and Australian validation of the spinal cord lesion-related coping strategies and emotional wellbeing questionnaires. *Spinal Cord* 2008;46:690–695. [PubMed: 18332886]
- Migliorini CE, Tonge BJ, Taleporos G. Spinal cord injury and mental health. *Australian and New Zealand Journal of Psychiatry* 2008;42:309–314. [PubMed: 18330773]
- Miller WC, Anton HA, Townson AF. Measurement properties of the CESD scale among individuals with spinal cord injury. *Spinal Cord* 2008;46:287–292. [PubMed: 17909558]
- Mitchell MC, Burns NR, Dorstyn DS. Screening for depression and anxiety in spinal cord injury with DASS-21. *Spinal Cord* 2008;46:547–551. [PubMed: 18071354]
- National Spinal Cord Injury Statistical Center. Annual Statistical Report. Birmingham, AL: University of Alabama, UAB; 2005.
- Oh SJ, Shin HI, Paik NJ, Yoo T, Ku JH. Depressive symptoms of patients using clean intermittent catheterization for neurogenic bladder secondary to spinal cord injury. *Spinal Cord* 2006;44:757–762. [PubMed: 16432529]
- Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction With Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment* 1991;57:149–161. [PubMed: 1920028]
- Pollard C, Kennedy P. A longitudinal analysis of emotional impact, coping strategies and post-traumatic psychological growth following spinal cord injury: a 10-year review. *British Journal of Health Psychology* 2007;12:347–362. [PubMed: 17640451]
- Richardson EJ, Richards JS. Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. *Rehabilitation Psychology* 2008;53(2):243–249.
- Taylor, G. Unpublished Dissertation. University of Minnesota; Minneapolis: 1967. Predicted versus actual response to spinal cord injury: A psychological study.
- Woolrich RA, Kennedy P, Tasiemski T. A preliminary psychometric evaluation of the Hospital Anxiety and Depression Scale (HADS) in 963 people living with a spinal cord injury. *Psychology Health and Medicine* 2006;11:80–90.
- Yardley JK, Rice RW. The relationship between mood and subjective well-being. *Social Indicators Research* 1991;24:101–111.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey MB, et al. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research* 1983;17:37–49. [PubMed: 7183759]



**Figure 1.**  
Receiver Operating Curve for OAHMQ on MDD



**Figure 2.**  
Receiver Operating Curve for OAHMQ on PHQ-9  $\geq 15$



**Figure 3.**  
Receiver Operating Curve for OAHMQ on PHQ-9  $\geq 10$

**Table 1**

Rates of depression symptom severity as classified by the PHQ-9 and OAHMQ broken down by race group.

Scale	Definition	Total	White	NonWhite
		% (n)		
<b>PHQ-9</b>				
No depressive symptoms	PHQ-9=0	17.7 (n=122)	15.4 (n=81)	25.0 (n=41)
Minimal depressive symptoms	PHQ-9=1 to 4	37.4 (n=258)	39.2 (n=206)	31.7 (n=52)
Mild depressive symptoms	PHQ-9=5 to 9	25.2 (n=174)	25.3 (n=133)	25.0 (n=41)
Moderate depressive symptoms	PHQ-9=10 to 14	10.4 (n=72)	10.5 (n=55)	10.4 (n=17)
Moderate/severe depressive symptoms	PHQ-9=15 to 19	5.4 (n=37)	5.3 (n=28)	5.5 (n=9)
Severe depressive symptoms	PHQ-9=20 to 27	3.9 (n=27)	4.4 (n=23)	2.4 (n=4)
Probable MDD	Reports at least 5/9 symptoms more than half the days in time period (must include either depressed mood or anhedonia)	10.7 (n=74)	10.8 (n=57)	10.4 (n=17)
<b>OAHMQ</b>				
Non-depressed	OAHMQ=0-5	55.6 (n=379)	57.5 (n=297)	49.7 (n=82)
Clinically significant symptomatology	OAHMQ=6-10	24.8 (n=169)	22.5 (n=116)	32.1 (n=53)
Probable major depression	OAHMQ=11-20	19.7 (n=134)	19.8 (n=102)	18.2 (n=30)

Table 2

Spearman Rank Correlations between various cut points of the OAHMQ, PHQ-9, and SWLS\*

	PHQ-9	MDD	PHQ-9 ≥15	PHQ-9 ≥10	OAHMQ	OAHMQ ≥6	OAHMQ ≥11	SWLS
PHQ-9	1.000							
MDD	0.530	1.000						
PHQ-9 ≥15	0.505	0.810	1.000					
PHQ-9 ≥10	0.692	0.700	0.645	1.000				
OAHMQ	0.781	0.477	0.440	0.621	1.000			
OAHMQ ≥6	0.678	0.389	0.356	0.556	0.863	1.000		
OAHMQ ≥11	0.610	0.605	0.527	0.693	0.690	0.553	1.000	
SWLS	-0.477	-0.335	-0.342	-0.408	-0.538	-0.485	-0.443	1.000

\* all correlations had a p-value &lt;.0001

**Table 3**

Mean value of SWLS by various cut points of the PHQ-9 and OAHMQ

	Mean (SD)	p-value*
MDD		
Yes	11.87 (6.49)	<.0001
No	21.67 (7.97)	
PHQ-9		
≥ 15	10.94 (5.96)	<.0001
< 15	21.62 (7.96)	
PHQ-9		
≥10	13.22 (7.15)	<.0001
<10	22.42 (7.65)	
OAHMQ		
≥ 11	12.46 (6.70)	<.0001
< 11	22.53 (7.61)	

\*  
from ANOVA F-test



**Table 4**

Frequencies of various cut points of the OAHMQ and PHQ-9

	<b>PHQ-9 MDD</b>		<b>Agreement</b>
	<b>Yes (n=74)</b>	<b>No (n=616)</b>	
	Column Percents		Kappa
PHQ-9			
≥15	77.03 (TP)	1.14 (FP)	0.81
0-14	22.97 (FN)	98.86 (TN)	
PHQ-9			
≥10	100.00 (TP)	10.06 (FP)	0.67
0-9	0.00 (FN)	89.94 (TN)	
OAHMQ			
≥6	100.00 (TP)	36.80 (FP)	0.26
0-5	0.00 (FN)	63.20 (TN)	
OAHMQ			
≥11	89.71 (TP)	11.24 (FP)	0.56
0-10	10.29 (FN)	88.76 (TN)	
	<b>PHQ-9 (≥15)</b>		
	<b>Yes (n=64)</b>	<b>No (n=626)</b>	
OAHMQ			
≥6	100.00 (TP)	37.86 (FP)	0.23
0-5	0.00 (FN)	62.14 (TN)	
OAHMQ			
≥11	86.21 (TP)	12.90 (FP)	0.48
0-10	13.79 (FN)	87.10 (TN)	
	<b>PHQ-9 (≥10)</b>		
	<b>Yes (n=136)</b>	<b>No (n=544)</b>	
OAHMQ			
≥6	99.22 (TP)	29.79 (FP)	0.48
0-5	0.78 (FN)	70.21 (TN)	
OAHMQ			
≥11	75.00 (TP)	5.88 (FP)	0.69
0-10	25.00 (FN)	94.12 (TN)	

TP=True positive which represents sensitivity, FP=False positive, FN=False negative, TN=True negative which represents specificity