Published in final edited form as:
Rehabil Psychol. 2014 May ; 59(2): 220–229. doi:10.1037/a0035919.

Comparing CESD-10, PHQ-9, and PROMIS Depression Instruments in Individuals with Multiple Sclerosis

Dagmar Amtmann,
Department of Rehabilitation Medicine, University of Washington, Seattle

Jiseon Kim,
Department of Rehabilitation Medicine, University of Washington, Seattle

Hyewon Chung,
Department of Education, Chungnam National University

Alyssa M. Bamer,
Department of Rehabilitation Medicine, University of Washington, Seattle

Robert L. Askew,
Department of Rehabilitation Medicine, University of Washington, Seattle

Salene Wu,
Department of Rehabilitation Medicine, University of Washington, Seattle

Karon F. Cook, and
Department of Medical Social Sciences, Northwestern University, Chicago

Kurt L. Johnson
Department of Rehabilitation Medicine, University of Washington, Seattle.

Abstract

Purpose—This study evaluated psychometric properties of the Patient Health Questionnaire-9 (PHQ-9), the Center for Epidemiological Studies Depression Scale-10 (CESD-10), and the eight-item PROMIS Depression Short Form (PROMIS-D-8; 8b short form) in a sample of individuals living with multiple sclerosis (MS).

Research Method—Data were collected by a self-reported mailed survey of a community sample of people living with MS (n=455). Factor structure, inter-item reliability, convergent/discriminant validity and assignment to categories of depression severity were examined.

Results—A one factor, confirmatory factor analytic model had adequate fit for all instruments. Scores on the depression scales were more highly correlated with one another than with scores on measures of pain, sleep disturbance, and fatigue. The CESD-10 categorized about 37% of participants as having significant depressive symptoms. At least moderate depression was

Correspondence concerning this article should be addressed to Dagmar Amtmann, Department of Rehabilitation Medicine, University of Washington, Box 354237, Seattle, WA 98195. dagmara@u.washington.edu.

The authors have no conflicts of interests related to the research in the manuscript. Portions of this paper were presented at the annual meeting of the consortium of multiple sclerosis centers, May 30 - June 2, 2012 at San Diego, CA.

indicated for 24% of participants by PHQ-9. PROMIS-D-8 identified 19% of participants as having at least moderate depressive symptoms and about 7% having at least moderately-severe depression. None of the examined scales had ceiling effects, but the PROMIS-D-8 had a floor effect.

Conclusions—Overall, scores on all three scales demonstrated essential unidimensionality and had acceptable inter-item reliability and convergent/discriminant validity. Researchers and clinicians can choose any of these scales to measure depressive symptoms in individuals living with MS. The PHQ-9 offers validated cut off scores for diagnosing clinical depression. The PROMIS-D-8 measure minimizes the impact of somatic features on the assessment of depression and allows for flexible administration, including Computerize Adaptive Testing (CAT). The CESD-10 measures two aspects of depression, depressed mood and lack of positive affect, while still providing an interpretable total score.

Keywords

depression; multiple sclerosis; CESD-10; PHQ-9; PROMIS

Multiple sclerosis (MS) is a chronic inflammatory disease of the brain and spinal cord. Individuals with MS are typically diagnosed in early to middle adulthood. Cognitive impairment and disruption of emotional and behavioral control and psychosocial functioning are often associated with MS (Bishop & Frain, 2011; Chiaravalloti & Deluca, 2002; Conway & Cohen, 2010; Feinstein, 2011; Halper et al., 2003). Common symptoms include fatigue, numbness, vision problems, dizziness and vertigo, pain, emotional changes, depressive symptoms, bowel and bladder dysfunction and spasticity (National Multiple Sclerosis Society, 2008). Depressive symptoms can be characterized by poor mood, losing interest in previously enjoyable experiences, fatigue, and feelings of worthlessness (Siegert & Abernethy, 2005). Studies related to MS and depressive symptoms suggest that people with MS experience significantly higher depressive symptoms than the general population (Chwastiak et al., 2002; Pattern, Beck, Williams, Barbui, & Metz, 2003; Patten, Metz, & Reimer, 2000; Rao, Huber, & Bomstein, 1992). One study found depression to be the most significant individual predictor of health distress in a sample of individuals with MS (White, White, & Russell, 2008). Furthermore, several MS studies have estimated that individuals with MS have a 37% to 54% risk of lifetime major depression that can dramatically affect their physical, social, and mental functioning (Chwastiak et al., 2002; Patten et al., 2003), with self-reported lifetime depression as high as 50% (Feinstein, 2011). The negative sequelae associated with depressive symptoms in MS include decreased perceived cognitive function (Maor, Olmer, & Mozes, 2001), increased fatigue (Koch, Mostert, Heerings, Uyttenboogaart, & De Keyser, 2009; Patten, Lavorato, & Metz, 2005) and sleep difficulties (Bamer, Johnson, Amtmann, & Kraft, 2010).

A number of self-reported instruments (scales or measures) have been used to screen for high depressive symptoms or major depressive disorder (MDD). The Patient Health Questionnaire-9 (PHQ-9) developed by Spitzer, Kroenke, and Williams (1999) is used to screen for MDD with items corresponding to the symptoms identified in the Diagnostic and Statistical Manual (American Psychiatric Association, 2000). The PHQ-9 also measures the severity of depressive symptoms and has been widely applied in medical settings (Kroenke,

Spitzer, Williams, & Löwe, 2010). Depressive symptoms also have been measured using the 20-item Center for Epidemiologic Studies Depression Scale (CESD-20) developed by Radloff (1977) to measure the severity of depressive symptoms in adults and adolescents. Unlike the PHQ-9, the CESD was originally constructed for use with the general community (Cole, Rabin, Smith, & Kaufman, 2004; Miller, Anton, & Townson, 2008). In addition, the 10-item version of the CESD (CESD-10) was developed to reduce respondent burden, (Andresen, Malmgren, Carter, & Patrick, 1994) and is well known for its quick administration and scoring (Sakakibara, Miller, Orenczuk, Wolfe, & SCIRE Research Team, 2009).

More recently, a depressive symptom item bank was developed by the National Institutes of Health's Patient Reported Outcome Measurement Information System (PROMIS) as one of many instruments to measure patient-reported outcomes relevant to a range of chronic diseases (Cella et al., 2010; Teresi et al., 2009). Both PHQ-9 and CESD-10 were developed within the Classical Test Theory (CTT) framework. PROMIS used an Item Response Theory (IRT) approach to develop item banks (as opposed to static instruments) to measure emotional distress, including depression (Pilkonis et al., 2011). The PROMIS item banks include a set of items calibrated to IRT that can be administered by Computerized Adaptive Testing (CAT) or by selecting a subset of items for use as fixed-length short forms. PROMIS developed several fixed length short forms. The 8-item version 1b (PROMIS-D-8) was used in this study. The PROMIS depression short form consists of eight items that were selected from the item bank based on CAT simulation results, item information, and content (Pilkonis et al., 2011). The IRT-based scoring of the PROMIS-D-8 was derived using all the items in the item bank, and as a result the PROMIS-D-8 is directly comparable to PROMIS Depression CAT scores and to the scores from different short forms, allowing meaningful comparisons across studies and populations. Another important advantage of the PROMIS-D-8 compared to other measures of depression, such as the PHQ-9 and CESD-10, is the lack of somatic symptoms that often overlap between depression and MS.

Because several measures of depressive symptoms are available, it is important to evaluate the psychometric strengths and weaknesses of each. Such information can help direct future research by identifying measures that are most suitable for a given purpose (e.g., epidemiological studies, clinical care). Instruments assessing depressive symptoms often sample items from multiple domains (e.g., mood, cognition, behavior, somatic symptoms) to capture a comprehensive set of manifest indicators of depression, and most instruments contain somatic items that could be related to disease rather than to depression in individuals with chronic illness and disability. Because of these differences, the factor structures of these instruments can differ from each other. The PROMIS-D item bank (and by extension the short form) was specifically developed to be unidimensional. While some studies of primary care and substance abuse samples have suggested the PHQ-9 is also unidimensional (Cameron, Crawford, Lawton, & Reid, 2008; Dum, Pickren, Sobell, & Sobell, 2008; Hansson, Chotai, Nordstöm, & Bodlund, 2009), studies of the measure in spinal cord injury samples have been mixed supporting both a 1-factor (Kalpakjian et al., 2009) and a 2-factor structure of affective and somatic symptoms (Richardson & Richards, 2008). The original, full length CESD-20 was found to have four factors of depressed mood, positive affect, somatic symptoms and interpersonal symptoms in people with MS (Verdier-Taillefer,

Gourlet, Fuhrer, & Alpérovitch, 2001). However, research on the factor structure of the CESD-10 for older adults has been mixed with some research support for a 2-factor structure of depressed mood and positive affect (Lee & Chokkanathan, 2008) and others suggesting a 3-factor structure of depressed mood, positive affect, and somatic symptoms (Cheng, Chan, & Fung, 2006).

Our literature review found no studies that examined the factor structure of the CESD-10 or the PHQ-9 in a sample of persons with MS. Differences in factor structure, and item content may indicate that different instruments measure different facets of depression. Determining which aspects of depression are most relevant to specific research protocols or clinical use is an important step in selecting among competing measures of depression.

The purpose of this study was to examine and compare psychometric properties of the PHQ-9, CESD-10, and PROMIS-D-8 in persons with MS and provide guidance to MS clinicians and researchers. We selected these instruments because they represent a diverse array of depression measures in medical populations, developed for different purposes and measuring different aspects of depression. The PHQ-9 corresponds to major depression and was developed for clinical use. The CES-D full version and short forms were developed for large epidemiological studies and sample a wide array of aspects of depression (mood, somatic). The PROMIS-D-8 excludes somatic symptoms and was developed for use with several medical populations. We evaluated unidimensionality, inter-item reliability, and convergent/discriminant validity, and assignment to depression severity categories based on scores. While factor structures of various measures for the same construct (such as depression) can differ, any instrument that provides a summary score needs to be sufficiently unidimensional. This is an assumption of both CTT and IRT based instruments (de Bonis, Lebeaux, de Boeck, Simon, & Pichot, 1991). This means that a set of items measures primarily the same construct. For instance, a summary score provided by a depression instrument, orders respondents on a continuum of depressive symptoms. Most commonly people with higher score have a higher level of depressive symptoms. If the item set is not sufficiently unidimensional and measures different constructs, the scores could not be ordered on one continuum and the summary score could not be meaningfully interpreted, because it could measure any of the dimensions or a mixture of dimensions. The assumption of unidimensionality must therefore be met by any instrument that provides a summary score.

Methods

Participants

Data for this study were collected as part of a longitudinal study of persons with MS. Research participants were recruited through the greater Washington chapter of the National Multiple Sclerosis Society (NMSS). Letters were sent to 7,806 persons from the NMSS mailing list. Eligibility criteria included being over the age of 18 and self-reporting having been diagnosed with MS by a physician. Of the 1,629 persons who responded, 1,597 were eligible and received a paper survey by mail. Reminder letters were sent to non-responders 3-6 weeks after the survey was mailed. There were 1,271 participants in the first survey and a random subset of participants (N=562) was invited to participate in the longitudinal study

that involved completing up to six surveys in four-month intervals. For the current study, data from the fifth time point, collected between June 2008 and December 2008, were used because all 3 instruments (PHQ-9, CESD-10, and PROMIS-D-8) were administered only at that time point. The Human Subjects Division at the University of Washington approved study procedures.

Instruments

Depression and depressive symptoms—The PHQ-9 (Spitzer et al, 1999) includes nine items with response options of 0 to 3 (0=Not at all; 1=Several days; 2=More than half the days; 3=Nearly every day). The time frame is “over the last 2 weeks”, and sum scores range from 0 to 27, with higher scores indicating more depressive symptoms. The nine items of the PHQ-9 correspond to the nine diagnostic criteria for a major depressive episode from the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000). The PHQ-9 is commonly used for screening and diagnosis, as well as selecting and monitoring treatment. PHQ-9 has also been used to measure depression in MS (Conway, Miller, O'Brien, & Cohen, 2012; Ferrando et al., 2007; Sjonnesen et al., 2012).

The CESD-10 (Andresen et al., 1994) is a short form that consists of ten items from the original 20 with response options of 0 to 3 [0=Rarely or none of the time (less than 1 day); 1=Some or a little of the time (1-2 days); 2=Occasionally or a moderate amount of time (3-4 days); 3=Most or all of the time (5-7 days)]. The time frame is “during the past week” and sum scores can range from 0 to 30, with higher scores indicating higher degrees of depressive symptoms. The CESD was designed to measure depressive experiences in the general population and includes items reflecting major dimensions of depression (Depressed Affect, Positive Affect, Somatic Symptoms/Retarded Activity and Interpersonal). Six items fall into Radloff's somatic symptoms/retarded activity grouping: (poor appetite, restless sleep, concentration, everything is an effort, could not get going, bothered by things that do not usually bother me). The CESD has also been used previously in MS research (Chwastiak et al., 2002; Patten et al., 2005).

The time frame for the PROMIS depression is “in the past seven days”, and the response options range from 1 to 5 (1=Never; 2=Rarely; 3=Sometimes; 4=Often; 5=Always). Scores are reported on the T-score metric [mean=50; standard deviation (SD) =10] that is centered on the general United States population mean in terms of age, gender and race/ethnicity (i.e., a score of 60 is one SD worse than the normative sample representative of the US general population) (Cella et al., 2010; Liu et al., 2010; Teresi et al., 2009). Of the 28 depression items in the PROMIS item bank, 17 are cognitive, 9 are affective, 1 is behavioral, and 1 reflects passive suicidal ideation. The item bank developers excluded behavioral and somatic items based on the results of psychometric analyses because somatic markers fit poorly (Pilkonis et al., 2011). In addition, it could be argued that the exclusion of most somatic features makes the bank more useful for assessing mood in people with chronic medical conditions and disabilities where physical symptoms may confound depression. The PROMIS short form used in this study consisted of eight items from the cognitive and affective categories that primarily focus on negative mood and negative views of the self. We know of only one previous study which used the PROMIS depression short form in

individuals with MS (Cook, Bamer, Amtmann, Molton, & Jensen, 2012). This study found no age or diagnosis related differential item function when comparing persons with MS to three other disability populations. Previous research has supported the test-retest reliability of each of the three measures considered in this study.

Although the CESD short form is less well studied, reported test-retest values for the CESD range from .40-.75 (time intervals from 2 weeks to 12 months) (Radloff, 1977; Vodermaier, Linden, & Siu, 2009). Reported test-retest values for the PHQ-9 range from .81 to .96 over a 7 day interval; the minimal clinically important difference for the PHQ-9 has been estimated as 5 score points (Lowe, Unutzer, Callahan, Perkins, & Kroenke, 2004). Test-retest reliability for the PROMIS-Depression measure has ranged from .66 to .78 across a 14 day interval (Narrow et al., 2013). Test-retest reliability could not be assessed in this study because of the cross-sectional nature of the data.

Other measures—The study also included measures of pain (PROMIS Pain Interference short form) (Amtmann et al., 2010), sleep (PROMIS Sleep Disturbance Short Form) (Pilkonis et al., 2011), and fatigue (Modified Fatigue Impact Scale; MFIS) (Fisk et al., 1994) to examine convergent/discriminant validity.

Analyses

Evaluation of dimensionality—Rather than exploring the dimensional structure of each instrument, our chief purpose was to examine whether the instruments are sufficiently unidimensional to ensure that a summary score is driven primarily by the construct of interest (i.e., depression) and is interpretable as such. A one-factor confirmatory factor analysis (CFA) was fit to examine dimensionality. Mplus software 6.1 (Muthén & Muthén, 1998-2010) was used to analyze the data with weighted least squares mean and variance adjusted (WLSMV) estimation. Goodness of fit was evaluated using χ^2, comparative fit index [CFI; (Bentler, 1980)], Tucker-Lewis index [TLI; (Tucker & Lewis, 1973)], and root mean square error of approximation [RMSEA; (Byrne, 1998; Steiger & Lind, 1980)]. CFI and TLI values above 0.95 are preferred (Hu & Bentler, 1999) and RMSEA values of less than 0.08 indicate adequate fit (Browne & Cudeck, 1993).

Inter-item reliability—To assess inter-item reliability, we calculated corrected item-total score correlations (i.e., each item score is correlated with the summed score based on all other items in the scale). Inter-item reliability is high if the items on an instrument measure the same construct. The corrected item-total score correlations based on Spearman's rank-order correlations were calculated using the raw scores of the PHQ-9, CESD-10, and PROMIS-D-8 (not the T-score). Corrected item-total score correlations greater than .40 are typically considered evidence of inter-item reliability (Everitt, 2002). These correlations were calculated using SAS software (version 9.3).

Discriminant/convergent validity—Validity evidence evaluates the degree to which a scale measures what it purports to measure. We reviewed convergent and discriminant validity evidence by examining the magnitude and direction of the correlations between the depression scores from the three instruments (i.e., convergent validity) and scores from

instruments designed to measure different constructs (discriminant validity). We expected high positive correlations among the scores from the three depression scales and moderate positive correlations with pain, sleep disturbance, and fatigue. We defined weak correlation as values between .20 and .40, moderate correlation as values between .41 and .70, and high correlation as values above .71 (Fountoulakis et al., 2007). Pearson correlations were used for the comparisons among scores on the PROMIS instruments, because all PROMIS instruments provide a T-score that is continuous. In all other cases, Spearman's rank-order correlation was calculated.

Severity categories—For the PHQ-9, we used the previously published cutoffs for severity categories and divided the MS sample into five categories of depressive symptoms (less than Mild /Mild/Moderate/Moderately severe/Severe) (Kroenke, Spitzer, & Williams, 2001). In addition, a cutoff score of 10 is often recommended to indicate probable MDD (Kroenke, Spitzer, & Williams, 2001). Although the CESD-10 is not intended to diagnose MDD, when it is used as a screening tool, a cutoff score greater than or equal to 10 has been suggested as indicating significant depressive symptoms (Andresen et al., 1994). We applied these cutoff scores in our study to compare the assignment by different instruments to categories based on severity of depression. Choi et al. (2012) developed a concordance table (i.e., score conversion table) between PHQ-9 and PROMIS depression measure using data from a large sample of the US general population. Thus, based on the conversion table, mild depression of PHQ-9 scores correspond to scores of [52.5, 58.6] on the PROMIS metric, moderate depression to scores of (58.6-64.7], moderately severe depression to scores of (64.7-70.3], and severe depression to scores of higher than 70.3. We applied those cutoff scores in our study to compare the assignment by different instruments to categories based on severity of depression.

Results

Participants

Participants with incomplete responses on the depression scales and demographic variables were not included in the study. The characteristics of the sample are described in Table 1. The average age of the sample was 53 years and the average years since MS diagnosis was 15. The mean PHQ-9 score was 6.6, suggesting that most participants had depressive symptoms below the moderate level. The average CESD-10 score was 8.5, suggesting that most individuals did not have clinically significant depressive symptoms [i.e., a cutoff score greater than or equal to 10 has been considered indicative of significant depressive symptoms (Andresen et al., 1994)]. The PROMIS depression mean was 50.1, indicating that mean depression levels in the MS sample were close to the US general population mean. Demographics for the sample were consistent with the distribution of MS in the general population with the exception of Caucasian race and education, which were higher than other MS samples. The sample was predominately female (83%) and white (91%). A total of 47% reported having a college or advanced degree. Almost 70% (n=316) were either married or living with a partner; 36% were employed. The most common self-reported type of MS (Bamer, Cetin, Amtmann, Bowen, & Johnson, 2007) in the sample was relapsing remitting (n=258, 58%), and 51% (n=231) obtained self-reported expanded disability status

scale (EDSS; Bowen, Gibbons, Gianas, & Kraft, 2001) mobility scores in the moderate level.

Analyses

Evaluation of unidimensionality—As shown in Table 2, fit indices from a one factor CFA for the PHQ-9, CESD-10, and PROMIS-D-8 were acceptable. CFI for all models was at or exceeded the recommended level of 0.95. TLIs for PHQ-9 and CESD-10 were just below the recommended level of 0.95 (0.94 and 0.93, respectively). TLI was above 0.95 for PROMIS-D-8. RMSEAs did not meet the recommended level for any of the measures.

Inter-item correlation—The range of the corrected item-total score correlation was from .35 to .67 for the PHQ-9, from .33 to .67 for the CESD-10, and from .75 to .84 for the PROMIS-D-8. The PHQ-9 had one item with a corrected item-total score correlation less than .40 (.35) (i.e., thoughts of being better off dead or of hurting yourself in some way). Most participants (90%) chose option 0 (not at all). In addition to this severe restriction of the range, the 9X9 PHQ-9 Spearman's rank-order item correlation matrix showed that this particular item produced the lowest correlation with six of the PHQ-9 items. The CESD-10 also had one item with a corrected item-total score correlation less than .40 (.33) (i.e., restless sleep). The 10×10 CESD-10 Spearman's rank-order item correlation matrix indicated that this item had the weakest association with the total score for seven of the CESD-10 items. Items whose correlation with the total score is less than .4 often do not contribute sufficiently to the total score (Amtmann et al., 2012; Everitt, 2002). The PROMIS-D-8 had no items with correlation less than .40.

Discriminant/convergent validity—As seen in Table 3, scores on depression scales were highly positively correlated with each other (.73-.85), moderately to highly correlated with fatigue scores (.55-.73), and moderately correlated with scores on measures of sleep disturbance (.39-.57) and pain (.47-.60). All correlations between measures were significant at the .01 alpha level.

Severity categories—Table 4 shows the distribution of scores falling within the PHQ-9, PROMIS-D-8 and CESD-10 severity categories. Based on the PHQ-9 severity categories, 24% of participants were classified as at least moderately depressed (score >=10), and 8% were categorized as at least moderately-severely depressed (score >=15). Based on the CESD-10 score of 10 or greater, 37% of participants were identified as having significant depression. Using the concordance table between PHQ-9 and PROMIS-D-8, 19% of participants were identified as having at least moderate depressive symptoms and about 7% having at least moderately-severe depression.

For every item on the PHQ-9, 6% (n=27) of the sample endorsed “Not at all” and 5.71% of the sample (n=26) endorsed the lowest category (i.e., “Rarely or none of the time”) for every item on the CESD-10 scale. Furthermore, 24.18% (n=110) of the sample endorsed “Never” for every item on the PROMIS-D-8. No participants answered “nearly every day” for every item on the PHQ-9, and “Most or all of the time” for every item of CESD-10 scale. Finally,

two responded in the highest category ("Always") for the every item on the PROMIS-D-8 scale suggesting negligible ceiling effects for all three scales.

Conclusions/Implications

The main objective of this study was to compare psychometric properties of the PHQ-9, CESD-10, and PROMIS-D-8 in a sample of persons with MS. Specifically, we assessed unidimensionality, inter-item reliability, convergent/discriminate validity, and score-based assignment to symptom severity categories. Unidimensionality is an important consideration for all self-reported scales, whether they were developed using CTT or an IRT approach, and the results of this study provided support for the unidimensionality of scores from the three measures. Scores on the PROMIS-D-8 and PHQ-9 have been found to be sufficiently unidimensional in several previous studies (Choi, Reise, Pilkonis, Hays, & Cella, 2010; Crane et al., 2010; Pilkonis et al., 2011). However, dimensionality analyses of CESD-10 scores have produced variable results, although depressed mood and positive affect factors consistently have been shown. Depressed mood and lack of positive affect are two highly related facets of the construct of depression (Watson et al., 1995). This study suggests that the CESD-10 scores are sufficiently unidimensional to warrant interpretation of total scores. The PROMIS-D-8 scores were highly unidimensional, most-likely because somatic items were removed from the item bank, and the scale was developed specifically to meet the unidimensionality assumption of IRT. Overall, these results support the continued use and interpretation of summary or total scores on these three measures in both research and clinical practice.

The corrected item-total score correlations were calculated as the internal reliability index. The PROMIS-D-8 had strong inter-item reliability with high associations between items and the total score. The correlations for all measures were greater than the criterion of .40, except for one item in both the PHQ-9 and CESD-10.While the psychometric functioning of these items is not optimal, they play an important role in screening for depression and therefore have been retained in the scales.

The results of the validity analyses supported the validity of the studied measures. Correlations among scores of all scales were in the expected direction and of expected magnitude. Consistent with previous findings (Bamer et al., 2010; Lobentanz et al., 2004; Newland, Fearing, Riley, & Neath, 2012), scores on all three measures were at least moderately correlated with fatigue, sleep disturbance, and pain interference. These correlations were somewhat higher than previously reported for other depressive symptom measures such as the Hospital Anxiety and Depression Scale and the Beck Depression Inventory (Motl & McAuley, 2009; Motl, Suh, & Weikert, 2010; Motl, Weikert, Suh, & Dlugonski, 2010; Newland et al., 2012).

Finally, the PHQ-9 identified about 24% of participants having at least moderate depressive symptoms and about 8% having at least moderately severe depression. The PROMISD-8 identified 19% of participants having at least moderate depressive symptoms and about 7% having at least moderately severe depression. The CESD-10 identified about 37% as significantly depressed (i.e., scoring equal to or higher than 10). The smaller proportion

classified with at least moderate depressive symptoms on the PROMIS-D-8 may have been due to the exclusion of somatic items. While the CESD-10 has three somatic items, the elevated proportion with clinically significant depressive symptoms compared to the other measures may have been due to the measure capturing mildly depressed participants as well as those with moderate or severe depression. The cutoff score was developed to detect depression and not necessarily moderate or more severe depression.

It has been recommended that floor and ceiling effects of instruments not exceed 15% (Hobart & Thompson, 2001). The PROMIS-D-8 also showed a floor effect (24%), suggesting it does not discriminate well among persons with very low levels of depressive symptoms. However, other measures showed minimal floor and ceiling effects.

Overall, the results of this study do not support the superiority of one instrument over the others in terms of psychometric properties. Scores on all scales showed similar characteristics and their correlations were high. The PHQ-9 was designed to incorporate all the DSM-IV depression elements and was mainly developed and tested for use with medical patients, not psychiatric patients or community residents, so it may be preferred if diagnosis or symptom monitoring is the main goal. Furthermore, it is widely regarded as easy to use by busy primary care practitioners (Bombardier, Richards, Krause, Tulsky, & Tate, 2004).

For epidemiological studies, where the goal is to estimate the severity of depressive symptoms, PHQ-9, CESD-10 and PROMIS-D-8 can all be used. Although evidence is not consistent, the physical symptoms associated with Major Depression may inflate severity scores in medical populations such as those with MS (Aikens et al., 1999; Cook et al., 2012; Mohr et al., 1997; Sjonnesen et al., 2012). Therefore, compared to the PHQ-9, the CESD short form and the PROMIS-D-8 may be preferable for research because they do not include as many (CESD) or any of the target somatic symptoms (PROMIS-D-8) that are diagnostic for both depression and MS. The CESD-10 may be particularly useful in studies that examine different aspects of depression such as depressed mood and positive affect. Treatment studies in particular may want to measure multiple aspects of depression to examine whether the treatment lessens depressed mood, increases positive affect or both. In contexts where cut-off scores are used to identify individuals for further evaluation, PROMIS cutoffs may be useful because compared to PHQ-9, the PROMIS-D-8 identified 5% fewer participants as having at least moderate depressive symptoms, potentially reducing the number of false positives. It is important to note that PROMIS-D-8 and CESD-10 do not ask about suicidal ideation, potentially limiting clinical utility. While this may enhance the psychometric performance of these measures, it may also limit their ability to screen for suicidal ideation. One option available to users of the PROMIS-D-8 and CESD-10 is to ask directly about suicidal ideation.

The availability of a minimal clinically important difference (MCID) estimate is an important consideration when selecting an instrument because it considerably enhances the interpretability of the instrument (Yost, Eton, Garcia, & Cella, 2011). The MCID for CESD-10 and PROMIS-D-8 could not be found in published literature. The MCID for the PHQ-9 has been estimated to be five points (Lowe et al., 2004). As a result, the PHQ-9 may

be preferable for treatment effectiveness studies where MCID can be used to evaluate whether changes in the scores are potentially meaningful or due to expected variation.

All three instruments evaluated in this study are brief and easy to administer and score. Though we evaluated a PROMIS short form, there are alternatives formats available within the PROMIS frameworks that are not available with the PHQ-9 and CESD-10. PROMIS items were developed using IRT (Cella et al., 2010) and as a result, allow for CAT administration, development of customized short forms targeted to specific populations or levels of depression, and more rigorous evaluation of bias via differential item functioning. CAT reduces respondent burden by reducing the number of items administered, as test items are selected and administered according to individual levels of symptoms, which simultaneously reduces assessment time and increases precision. Furthermore, CAT has been found to increase respondents' motivation, because items are adaptive to each patient's individual level of symptom burden and therefore are more relevant (Gardner et al., 2004; Gibbons et al., 2008). Pilkonis et al. (2011) found that just a few items (in most cases, four to six) need to be administered when using CAT. An additional advantage of the PROMIS measures is that scores are centered on the United States general population mean, thereby assisting in interpretation of scores (Pilkonis et al., 2011); algorithms have also been developed to translate scores from the PHQ-9 to the PROMIS metric to allow researchers to maintain continuity with previous research. (Choi et al., 2012).

This study is not without limitations. This convenience sample was relatively well-educated and largely Caucasian and is not representative of the population of individuals living with MS. Other limitations included the lack of clinical confirmation of the self-reported MS diagnosis by a physicians and the lack of a gold-standard assessment of depression, such as a structured clinical interview. The lack of a criterion measure prevented evaluation of the instruments' utility as a screen for MDD. Therefore cautious interpretation and generalization of findings may be necessary until the findings are replicated in other samples.

Future research is needed that investigates the cut-score based severity classifications against a clinical standard, such as the SCID (First, 2005). Data used in this study could not be used to investigate the impact of trans-diagnostic symptoms on accuracy of screening for MDD because no instrument that could be used as a clinical standard was administered. Such research would be useful in discriminating among the instruments with respect to their use for screening and referral for services. Additional research could also address the overlap of somatic symptoms between MS and depression, possibly by examining alternate criteria for major depression (Cavanaugh, 1995; Endicott, 1984). One such approach could include assessment of Differential Item Functioning (DIF) within the IRT framework. While DIF is well suited for investigating the impact of trans-diagnostic symptoms on screening for MDD in people with MS, it requires data from a reference sample for all three instruments; such data were not available for this study. Lastly, additional research is needed to establish the minimal clinically important differences for the CESD-10 and the PROMIS-D-8 in MS.

In summary, the PHQ-9, CESD-10, and PROMIS-D-8 demonstrated comparable psychometric properties in a sample of individuals living with MS. While all can be useful in MS research and clinical practice, the PHQ-9 may be preferable in clinical practice, because of its validated cut-off scores and similarity to DSM-IV symptoms. While the CESD-10 and PROMISD-8 may both be preferable for large-scale research studies, researchers and clinicians may find specific features of the PROMIS measures (e.g. multiple forms of administration, customizable content, and the US general population norms) to be useful in their research and clinical practice.

Acknowledgments

The content of this manuscript was developed under the grants from the Department of Education, NIDRR grant number H133B080025 and H133P120002, and the National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases (Grant #U01AR052171, Dagmar Amtmann, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Department of Education, and you should not assume endorsement by the Federal Government.

References

Aikens JE, Reinecke MA, Pliskin NH, Fischer JS, Wiebe JS, McCracken LM, Taylor JL. Assessing depressive symptoms in multiple sclerosis: is it necessary to omit items from the original Beck Depression Inventory? Journal of Behavioral Medicine. 1999; 22(2):127–142. [PubMed: 10374139]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4th ed., text revision ed.. American Psychiatric Association; Washington, DC: 2000.

Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, Lai JS. Development of a PROMIS item bank to measure pain interference. [Research Support, N.I.H., Extramural Validation Studies]. Pain. 2010; 150(1):173–182. doi: 10.1016/j.pain.2010.04.025. [PubMed: 20554116]

Amtmann D, Bamer AM, Noonan V, Lang N, Kim J, Cook KF. Comparison of the psychometric properties of two fatigue scales in multiple sclerosis. Rehabilitaiton Psychology. 2012; 57(2):159–166. doi: 10.1037/a0027890.

Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). American Journal of Preventive Medicine. 1994; 10(2):77–84. [PubMed: 8037935]

Bamer AM, Johnson KL, Amtmann DA, Kraft GH. Beyond fatigue: Assessing variables associated with sleep problems and use of sleep medications in multiple sclerosis. Clinical Epidemiolology. 2010; 2010(2):99–106. doi: 10.2147/CLEP.S10425.

Bamer AM, Cetin K, Amtmann D, Bowen JD, Johnson KL. Comparing a self report questionnaire with physician assessment for determining multiple sclerosis clinical disease course: a validation study. Multiple Sclerosis. 2007; 13(8):1033–7. [PubMed: 17468438]

Bentler PM. Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology. 1980; 31(1):419–456. doi: 10.1146/annurev.ps.31.020180.002223.

Bishop M, Frain MP. The Multiple Sclerosis Self-Management Scale: Revision and psychometric analysis. Rehabilitation Psychology. 2011; 56(2):150–159. doi: 2011-09991-010. [PubMed: 21574735]

Bombardier CH, Richards JS, Krause JS, Tulsky 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. doi: S0003999304010846. [PubMed: 15520969]

Bowen J, Gibbons L, Gianas A, Kraft GH. Self-administered Expanded Disability Status Scale with functional system scores correlates well with a physician-administered test. Multiple Sclerosis. 2001; 7(3):201–206. [PubMed: 11475445]

Browne, MW.; Cudeck, R. Alternative Ways of Assessing Model Fit. In: Bollen, KA.; Long, S., editors. Testing Structural Equation Models. Sage Publications; Newbury Park, CA: 1993.

Byrne, BM. Structural equation modeling with LISREL, PRELIS, and SIMPLIS. Lawrence Erlbaum; Hillsdale, NJ: 1998.

Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. British Journal of General Practice. 2008; 58(546): 32–36. doi: 10.3399/bjgp08X263794. [PubMed: 18186994]

Cavanaugh S. Depression in the medically ill: Critical issues in diagnostic assessment. Psychosomatics. 1995; 36:48–59. doi:10.1016/S0033-3182(95)71707-8. [PubMed: 7871134]

Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Hays R. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. Journal of Clinical Epidemiology. 2010; 63(11): 1179–1194. doi: 10.1016/j.jclinepi.2010.04.011. [PubMed: 20685078]

Cheng ST, Chan AC, Fung HH. Factorial structure of a short version of the Center for Epidemiologic Studies Depression Scale. International Journal of Geriatric Psychiatry. 2006; 21(4):333–336. doi: 10.1002/gps.1467. [PubMed: 16570325]

Chiaravalloti ND, Deluca J. Self-generation as a means of maximizing learning in multiple sclerosis: an application of the generation effect. Archives Physical Medicine and Rehabilitation. 2002; 83(8):1070–1079. doi: 10.1053/apmr.2002.33729.

Choi SW, Podrabsky T, McKinney N, Schalet BD, Cook KF, Cella D. PROSetta Stone® Analysis Report: a Rosetta Stone for Patient Reported Outcomes. 2012; 1 Retrived from http://www.prosettastone.org/AnalysisReport/Documents/PROsettaStoneAnalysisReportVol1.pdf.

Choi SW, Reise SP, Pilkonis PA, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. Quality of Life Research. 2010; 19(1):125–136. doi: 10.1007/s11136-009-9560-5. [PubMed: 19941077]

Chwastiak L, Ehde DM, Gibbons LE, Sullivan M, Bowen JD, Kraft GH. Depressive symptoms and severity of illness in multiple sclerosis: epidemiologic study of a large community sample. American Journal of Psychiatry. 2002; 159(11):1862–1868. doi:10.1176/appi.ajp.159.11.1862. [PubMed: 12411220]

Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychological Assessment. 2004; 16(4):360–372. doi: 2004-21195-003. [PubMed: 15584795]

Conway D, Cohen JA. Emerging oral therapies in multiple sclerosis. Current Neurology and Neuroscience Reports. 2010; 10(5):381–388. doi: 10.1007/s11910-010-0125-3. [PubMed: 20549393]

Conway DS, Miller DM, O'Brien RG, Cohen JA. Long term benefit of multiple sclerosis treatment: an investigation using a novel data collection technique. Multiple Sclerosis. 2012; 18(11):1617–1624. doi: 10.1177/1352458512449681. [PubMed: 22653659]

Cook KF, Bamer AM, Amtmann D, Molton IR, Jensen MP. Six patient-reported outcome measurement information system short form measures have negligible age- or diagnosis-related differential item functioning in individuals with disabilities. Archives of Physical Medicine and Rehabilitation. 2012; 93(7):1289–1291. doi: 10.1016/j.apmr.2011.11.022. [PubMed: 22386213]

Crane PK, Gibbons LE, Willig JH, Mugavero MJ, Lawrence ST, Schumacher JE, Crane HM. Measuring depression levels in HIV-infected patients as part of routine clinical care using the nine-item Patient Health Questionnaire (PHQ-9). AIDS Care. 2010; 22(7):874–885. doi: 10.1080/09540120903483034. [PubMed: 20635252]

de Bonis M, Lebeaux MO, de Boeck P, Simon M, Pichot P. Measuring the severity of depression through a self-report inventory. A comparison of logistic, factorial and implicit models. Journal of Affective Disord. 1991; 22(1-2):55–64. doi:10.1016/01650327(91)90084-6.

Dum M, Pickren J, Sobell LC, Sobell MB. Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addictive Behaviors. 2008; 33(2):381–387. doi: 10.1016/j.addbeh.2007.09.017. [PubMed: 17964079]

Endicott J. Measurement of depression in patients with cancer. Cancer. 1984; 53(10 Suppl):2243–2249. [PubMed: 6704912]

Everitt, BS. The cambridge dictionary of statistics. 2nd ed.. Cambridge University Press; New York: 2002.

Feinstein A. Multiple sclerosis and depression. Multiple Sclerosis Journal. 2011; 17(11):1276–1281. doi: 10.1177/1352458511417835. [PubMed: 22058085]

Ferrando SJ, Samton J, Mor N, Nicora S, Findler M, Apatoff B. Patient Health Questionnaire-9 to Screen for Depression in Outpatients With Multiple Sclerosis. International Journal of MS Care. 2007; 9(3):99–103. doi: http://dx.doi.org/10.7224/1537-2073-9.3.99.

First M. Clinical utility: a prerequisite for the adoption of a dimensional approach in. Journal of Abnormal Psychology. 2005; 114(4):560–564. [PubMed: 16351379]

Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clinical Infectious Diseases. 1994; 18(Suppl 1):S79–83. [PubMed: 8148458]

Fountoulakis KN, Bech P, Panagiotidis P, Siamouli M, Kantartzis S, Papadopoulou A, St Kaprinis G. Comparison of depressive indices: reliability, validity, relationship to anxiety and personality and the role of age and life events. Journal of Affective Disorders. 2007; 97(1-3):187–195. doi: 10.1016/j.jad.2006.06.015. [PubMed: 16844229]

Gardner W, Shear K, Kelleher KJ, Pajer KA, Mammen O, Buysse D, Frank E. Computerized adaptive measurement of depression: a simulation study. BMC Psychiatry. 2004; 4:13. doi: 10.1186/1471-244X-4-13. [PubMed: 15132755]

Gibbons RD, Weiss DJ, Kupfer DJ, Frank E, Fagiolini A, Grochocinski VJ, Immekus JC. Using computerized adaptive testing to reduce the burden of mental health assessment. Psychiatric services. 2008; 59(4):361. doi: 10.1176/appi.ps.59.4.361. [PubMed: 18378832]

Halper J, Kennedy P, Miller CM, Morgante L, Namey M, Ross AP. Rethinking cognitive function in multiple sclerosis: a nursing perspective. Journal of Neuroscience Nursing. 2003; 35(2):70–81. [PubMed: 12795033]

Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. British Journal of General Practice. 2009; 59(566):e283–288. doi: 10.3399/bjgp09X454070. [PubMed: 19761655]

Hobart, J.; Thompson, A. Assessment measures and clinical scales.. In: Guiloff, R., editor. Clinical Trials in Neurology. Springer-Verlag; London: 2001. p. 17-28.

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999; 6(1):1–55. doi: 10.1080/10705519909540118.

Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG. Patient health Questionnaire-9 in spinal cord injury: an examination of factor structure as related to gender. Journal of Spinal Cord Medicine. 2009; 32(2):147–156. [PubMed: 19569462]

Koch M, Mostert J, Heerings M, Uyttenboogaart M, De Keyser J. Fatigue, depression and disability accumulation in multiple sclerosis: a cross-sectional study. European Journal of Neurology. 2009; 16(3):348–352. doi: 10.1111/j.1468-1331.2008.02432.x. [PubMed: 19490071]

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. doi: jgi01114. [PubMed: 11556941]

Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. General Hospital Psychiatry. 2010; 32(4): 345–359. doi: 10.1016/j.genhosppsych.2010.03.006. [PubMed: 20633738]

Lee AE, Chokkanathan S. Factor structure of the 10-item CES-D scale among community dwelling older adults in Singapore. International Journal of Geriatrric Psychiatry. 2008; 23(6):592–597. doi: 10.1002/gps.1944.

Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, Hays RD. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. Journal of Clinical Epidemiology. 2010; 63(11):1169–1178. doi: 10.1016/j.jclinepi. 2009.11.021S0895-4356(10)00172-1. [PubMed: 20688473]

Lobentanz IS, Asenbaum S, Vass K, Sauter C, Klösch G, Kollegger H, Zeitlhofer J. Factors influencing quality of life in multiple sclerosis patients: disability, depressive mood, fatigue and sleep quality. Acta Neurologica Scandinavica. 2004; 110(1):6–13. doi: 10.1111/j. 1600-0404.2004.00257.x. [PubMed: 15180801]

Lowe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Medical Care. 2004; 42(12):1194–1201. doi: 00005650-200412000-00006. [PubMed: 15550799]

Maor Y, Olmer L, Mozes B. The relation between objective and subjective impairment in cognitive function among multiple sclerosis patients-the role of depression. Multiple Sclerosis. 2001; 7(2): 131–135. [PubMed: 11424633]

Miller WC, Anton HA, Townson AF. Measurement properties of the CESD scale among individuals with spinal cord injury. Spinal Cord. 2008; 46(4):287–292. doi: 10.1038/sj.sc.3102127. [PubMed: 17909558]

Mohr DC, Goodkin DE, Likosky W, Beutler L, Gatto N, Langan MK. Identification of Beck Depression Inventory items related to multiple sclerosis. Journal of Behavioral Medicine. 1997; 20(4):407–414. doi:10.1023/A:1025573315492. [PubMed: 9298438]

Motl RW, McAuley E. Symptom cluster as a predictor of physical activity in multiple sclerosis: preliminary evidence. Journal of Pain and Symptom Management. 2009; 38(2):270–280. doi: 10.1016/j.jpainsymman.2008.08.004. [PubMed: 19329276]

Motl RW, Suh Y, Weikert M. Symptom cluster and quality of life in multiple sclerosis. Journal of Pain and Symptom Management. 2010; 39(6):1025–1032. doi: 10.1016/j.jpainsymman.2009.11.312. [PubMed: 20434872]

Motl RW, Weikert M, Suh Y, Dlugonski D. Symptom cluster and physical activity in relapsing-remitting multiple sclerosis. Research in Nursing and Health. 2010; 33(5):398–412. doi: 10.1002/nur.20396. [PubMed: 20725947]

Muthén, LK.; Muthén, BO. Mplus User's Guide. 6th ed.. Muthén & Muthén; Los Angeles, CA: 1998-2010.

Narrow WE, Clarke DE, Kuramoto SJ, Kraemer HC, Kupfer DJ, Greiner L, Regier DA. DSM-5 field trials in the United States and Canada, Part III: development and reliability testing of a cross-cutting symptom assessment for DSM-5. American Journal of Psychiatry. 2013; 170(1):71–82. doi: 10.1176/appi.ajp.2012.12071000. [PubMed: 23111499]

National Multiple Sclerosis Society. Clinical Study Measures: Mental Health Inventory. 2008. Retrieved June 27, 2012, from http://www.nationalmssociety.org/for-professionals/researchers/clinical-study-measures/mhi/index.aspx

Newland PK, Fearing A, Riley M, Neath A. Symptom clusters in women with relapsing-remitting multiple sclerosis. Journal of Neuroscience Nursing. 2012; 44(2):66–71. doi: 10.1097/JNN.0b013e3182478cba. [PubMed: 22367268]

Patten SB, Beck CA, Williams JV, Barbui C, Metz LM. Major depression in multiple sclerosis: a population-based perspective. Neurology. 2003; 61(11):1524–1527. [PubMed: 14663036]

Patten SB, Lavorato DH, Metz LM. Clinical correlates of CES-D depressive symptom ratings in an MS population. General Hospital Psychiatry. 2005; 27(6):439–445. doi: 10.1016/j.genhosppsych.2005.06.010. [PubMed: 16271659]

Patten SB, Metz LM, Reimer MA. Biopsychosocial correlates of lifetime major depression in a multiple sclerosis population. Multiple Sclerosis. 2000; 6(2):115–120. doi: 10.1177/135245850000600210. [PubMed: 10773857]

Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, Group PC. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment. 2011; 18(3):263–283. doi: 10.1177/1073191111411667. [PubMed: 21697139]

Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1(3):385–401. doi: 10.1177/014662167700100306.

Rao SM, Huber SJ, Bornstein RA. Emotional changes with multiple sclerosis and Parkinson's disease. Journal of Consulting and Clinical Psychology. 1992; 60(3):369–78. [PubMed: 1619091]

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. doi: 10.1037/0090-5550.53.2.243.

Sakakibara BM, Miller WC, Orenczuk SG, Wolfe DL, SCIRE Research Team Team. A systematic review of depression and anxiety measures used with individuals with spinal cord injury. Spinal Cord. 2009; 47(12):841–851. doi: 10.1038/sc.2009.93. [PubMed: 19621021]

Siegert RJ, Abernethy DA. Depression in multiple sclerosis: a review. Journal of Neurology, Neurosurgery, and Psychiatry. 2005; 76(4):469–475. doi:10.1136/jnnp.2004.054635.

Sjonnesen K, Berzins S, Fiest KM, M Bulloch AG, Metz LM, Thombs BD, Patten SB. Evaluation of the 9-item Patient Health Questionnaire (PHQ-9) as an assessment instrument for symptoms of depression in patients with multiple sclerosis. Postgradraduate Medicine. 2012; 124(5):69–77. doi: 10.3810/pgm.2012.09.2595.

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. The Journal of the American Medical Association. 1999; 282(18):1737–1744.

Steiger, JH.; Lind, JC. Statistically-based tests for the number of common factors.. Paper presented at the Annual spring meeting of the Psychometric Society; Iowa City, IA. 1980.

Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, Cella D. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. Psychol Sci Q. 2009; 51(2):148–180. [PubMed: 20336180]

Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973; 38:1–10. DOI:10.1007/BF02291170.

Verdier-Taillefer MH, Gourlet V, Fuhrer R, Alpérovitch A. Psychometric properties of the Center for Epidemiologic Studies-Depression scale in multiple sclerosis. Neuroepidemiology. 2001; 20(4): 262–267. doi:10.1159/000054800. [PubMed: 11684903]

Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. Journal of National Cancer Institute. 2009; 101(21):1464–1488. doi: 10.1093/jnci/djp336.

Watson D, Clark LA, Weber K, Assenheimer JS, Strauss ME, Mccormick RA. Testing a Tripartite Model: 1. Evaluating the Convergent and Discriminant Validity of Anxiety and Depression Symptom Scales. Journal of Abnormal Psychology. 1995; 104(1):3–14. [PubMed: 7897050]

White CP, White MB, Russell CS. Invisible and visible symptoms of multiple sclerosis: which are more predictive of health distress? Journal of Neuroscience Nursing. 2008; 40(2):85–95. 102. doi: 10.1097/01376517-200804000-00007. [PubMed: 18481738]

Yost JK, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six patient-reported outcomes measurement information system-cancer scales in advanced-stage cancer patients. Journal of Clinical Epidemiology. 2011; 64:507–516. doi: 10.1016/j.jclinepi.2010.11.018. [PubMed: 21447427]

Impact

- A comprehensive psychometric evaluation of common depression measures in people with Multiple Sclerosis (MS) has not yet been reported. This study provides information and empirical support that can help researchers and clinicians select an appropriate instrument for measuring depressive symptoms in MS.
- This study found that the Patient Health Questionnaire 9 (PHQ-9), Center for Epidemiological Studies-Depression short form (CESD-10) and the PROMIS 8-item depression scale (PROMIS-D-8) had sufficient unidimensionality and provided evidence for reliability and validity of all three scores in people with MS.
- The PHQ-9 may be particularly well suited for clinical practice while the CESD-10 and PROMIS-D-8 may be well suited for research studies.

Table 1

Demographic Characteristics of a Sample of Individuals with Multiple Sclerosis at Time Point 5

	MS time point 5 (n=455) n (%) mean ± SD	
Age	52.9±10.8	
Years since MS diagnosis	14.5±10.0	
PHQ-9	6.6±5.2	
CESD-10	8.5±6.3	
PROMIS-D-8	50.2±9.9	
Pain interference (n=363)	57.4±8.2	
Modified fatigue impact scale	38.9±19.1	
Gender		
Male	78	17.1%
Female	377	82.9%
Ethnicity		
White	416	91.4%
Non-white	39	8.6%
Marriage Status		
Never-married	21	4.6%
Married/ Living with partner in committed relationship	316	69.5%
Separated/Divorced/Widowed	118	25.9%
Education		
Less than high school grad	2	0.4%
High school grad/GED	61	13.4%
Vocational or technical school	40	8.8%
Some college/Technical degree/AA	136	29.9%
College degree (BA/BS)	134	29.5%
Advanced degree (MA, PHD, MD)	82	18.0%
MS Type (self-reported)		
Relapsing remitting	258	56.7%
Other types	188	41.3%
Missing	9	2.0%
Employment		
Unemployed	290	63.7%
Employed	165	36.3%
Income		
Less than $25,000	76	16.7%
$25,000-$40,000	67	14.7%
$41,000-$55,000	50	11.0%
$56,000-$70,000	65	14.3%
$71,000-$85,000	40	8.8%
$86,000-$100,000	51	11.2%
Greater than $100,000	77	16.9%

	MS time point 5 (n=455) n (%) mean ± SD	
Decline to answer	29	6.4%

Table 2

Model Fit for a One Factor CFA Analysis for Depression Scales

	χ^2	DF	CFI	TLI	RMSEA (90% C.I.)
PHQ-9	174.38	27	0.95	0.94	0.11 (0.09, 0.13)
CESD-10	207.03	35	0.95	0.93	0.10 (0.09, 0.12)
PROMIS-D-8	300.26	20	0.99	0.98	0.18 (0.16, 0.19)

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; C.I. = confidence interval;

Table 3

Correlations between Depression Scales and Other Measures.

Measure	PHQ-9	CESD-10	PROMIS-D-8	MFIS	PROMIS-Sleep Disturbance
PHQ-9					
CESD-10	0.85				
PROMIS-D-8	0.73	0.80			
MFIS	0.73	0.71	0.55		
PROMIS-Sleep Disturbance	0.57	0.56	0.39[a]	0.36	
PROMIS-Pain interference	0.60	0.55	0.47[a]	0.69	0.34[a]

Note: MFIS= Modified Fatigue Impact Scale. PROMIS scores are based on T-scores.

[a]Pearson correlation, otherwise Spearman rank-order correlation; all correlations are significant at p<.01 (n of pain=363; otherwise n=455).

Table 4

Distribution of Participants across the PHQ-9, CESD-10, and PROMIS-D-8 Severity Ratings.

PHQ-9	n(%)
Less than mild or minimal <5	193 (42.4%)
Mild [5, 9]	151 (33.2%)
Moderate [10, 14]	74 (16.3%)
Moderately Severe [15, 19]	18 (3.9%)
Severe >19	19 (4.2%)

CESD-10	n(%)
Significant Depression Symptom >=10	169 (37.1%)

PROMIS-D-8	n(%)
Less than mild or minimal <52.5	272 (59.8%)
Mild [52.5, 58.6]	95 (20.9%)
Moderate (58.6, 64.7]	56 (12.3%)
Moderately Severe (64.7, 70.3]	20 (4.4%)
Severe >70.3	12 (2.6%)