Published in final edited form as:

Am J Geriatr Psychiatry. 2015 November ; 23(11): 1134–1143. doi:10.1016/j.jagp.2015.04.004.

Comparing PROMIS to Legacy Depression Measures in a Community Sample of Older Adults with Varying Levels of Cognitive Functioning

Jennifer B. Levin, PhD^{1,2}, Michelle E. Aebi, BA¹, Kathleen Smyth, PhD³, Curtis Tatsuoka, PhD^{2,3,4}, Johnny Sams, MA, MBA^{1,2}, Thomas Scheideman, MD¹, and Martha Sajatovic, MD^{1,2,4}

¹ Department of Psychiatry, Case Western Reserve University School of Medicine, Cleveland, Ohio.

² Neurological and Behavioral Outcomes Center, University Hospitals Case Medical Center, Cleveland, Ohio.

³ Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, Ohio.

⁴ Department of Neurology, Case Western Reserve University School of Medicine, Cleveland, Ohio.

Abstract

Objectives—This study evaluated the utility of Patient-Reported Outcomes Measure Information System Depression Scale (PROMIS-8a) compared to selected “Legacy” depression scales including Montgomery-Asberg Depression Rating Scale (MADRS), Geriatric Depression Scale (GDS), and GDS-Short Form (GDS-SF). Additionally, the measures’ properties were assessed across levels of cognitive functioning.

Methods—This cross-sectional analysis was extracted from a prospective cohort study. PROMIS-8a and Legacy depression measures were administered to individuals aged 70 grouped by cognitive status based on the Saint Louis University Mental Status Examination. McNemar tests were run to determine if measures categorized the absence or presence of depression differently and item analysis evaluated classification discrepancies.

Results—Sample mean age was 78, the majority was female (71%), Caucasian (79%), with at least a high-school education (89%). The percentage of individuals with at least mild depression was similar across measures (20.7% PROMIS-8a; 19.0% MADRS; 17.9% GDS; 13.9% GDS-SF).

Address Correspondence to: Jennifer B. Levin, Ph.D., Department of Psychiatry, University Hospitals Case Medical Center, 10524 Euclid Avenue, Cleveland, Ohio, 44106 Tel: 216-844-5057, Fax: 216-844-1703, Jennifer.levin@uhhospitals.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Previous Presentation:

This paper was presented at the International Society for CNS Clinical Trials and Methodology (ISCTM) Conference in Boston, MA, on October 6-8, 2014.

PROMIS-8a total score correlated moderately with MADRS ($r=.56$, $df=295$, $p<.01$), GDS ($r=.68$, $df=291$, $p<.01$), and GDS-SF ($r=.60$, $df=291$, $p<.01$) and predictive validity of the measures was similar. There were no significant mean differences on depression measures by cognitive status.

Conclusions—While all measures identified a similar percent of depressed individuals, the classification differed by measure. Item analysis showed that PROMIS-8a was more likely to identify feelings of dysphoria, and the MADRS and GDS were more likely to identify physiological aspects of depression. Given the brevity and ease of administration of the PROMIS-8a, it appears to be a useful depression screen for community-dwelling older adults.

Objectives

Depression is common among older adults (1) and is a risk factor for cognitive impairment (2). In the general population, between 3-26% of older adults have significant depression (3), a condition rated as the fourth leading cause of disease burden (4). While public campaigns to recognize and address depression have been helpful, depression is still under-detected in older adults living in the community (5). Furthermore, depressed individuals are at increased risk for developing dementia (6). Cognitive impairment is also common among older adults with 14% of adults over 70 years of age and 37% of those over 90 meeting criteria for dementia (3). Later-life depression predicts worse functional ability in older adults, which is especially true for those who are persistently, rather than intermittently, depressed (7).

Recently, the National Institute of Health (NIH) funded the Patient-Reported Outcomes Measurement Information System (PROMIS) which measures seven common health domains including emotional distress (depression), anxiety, pain, fatigue, sleep disturbance, physical functioning, and social participation (8). The NIH describes PROMIS as a “psychometrically validated, dynamic system to measure PROs [Patient-Reported Outcomes] efficiently in study participants with a wide range of chronic diseases and demographic characteristics” (8). The PROMIS depression scale was specifically developed to be unidimensional and targets a maximum reading level of the sixth grade. The PROMIS is also purported to be a flexible measure of depression as it is comprised of various possible short-forms. The PROMIS depression battery is self-rated and generally takes under 5 minutes to complete (8).

Various self-report and clinician-rated “Legacy” measures have been developed to assess depression in adults across the age-span. The Montgomery-Asberg Depression Rating Scale (MADRS) was developed to detect change in depressive symptomatology in adults as a function of antidepressant treatment and is now often used by researchers to evaluate depressive severity over time (9, 10). Administering the MADRS requires a well-trained rater and generally takes approximately 15-20 minutes. The MADRS measures various domains of depression in an attempt to create an overarching picture of symptom indicators with a relative lack of emphasis on somatic symptoms (11).

The Geriatric Depression Scale (GDS) is a depression screening measure for older adults. This 30-item self-rated measure was designed to alleviate the reliance on questions that address somatic complaints that may confound the detection of depression in older adults

(12, 13). The GDS can generally be completed by a patient in 10 minutes or less. Similar to the MADRS, the GDS has multiple domains that measure various aspects of depression (11). The GDS Short Form (GDS-SF) consists of 15 items extracted from the GDS. Lelito et al. (2001) found the short form to be as effective as the original form in identifying depression (14). Studies show a high correlation between the long and short form ($r = 0.91$, $p < .01$) (12). According to Friedman et al. (2005), the GDS-SF has high internal consistency (Cronbach $\alpha = .75$) and is comprised of two domains: depression and positive affect (15). The GDS-SF can be completed in about half the time as the full-length GDS.

Despite the various available tools to screen and diagnose depression in the general population, few studies have compared the strengths, weaknesses and psychometric properties of these scales in an older adult population with varying levels of cognitive functioning. The aim of this baseline analysis extracted from a larger prospective study was to compare the PROMIS-8a depression scale to Legacy measures of depression (MADRS, GDS, GDS-SF) in individuals age 70 or over. Additionally, this study aimed to examine the performance of the four depression measures in older adults with various levels of cognitive functioning. We hypothesized that there would be no difference between the PROMIS-8a and the Legacy depression measures' ability to identify older adults at risk for depression and that there would be no difference in the measures' ability to detect mild depression as it relates to different levels of cognitive functioning. The MADRS was used as the gold standard for identifying who was at risk for at least mild depression.

Methods

Subjects

A sample of 304 study participants were recruited directly from various community-dwelling residential settings (see Table 1 for housing status) in northeast Ohio as well as from an outpatient clinical practice specializing in neurology and geriatrics. A minority of the subjects (3%) were recruited during an inpatient stay on a geropsychiatric unit of an academic medical center. On average, an inpatient stay on this unit ranges from 5-8 days. Inclusion criteria were purposely broad in order to represent individuals that comprise community-dwelling older adult populations. Eligible participants were: 1) age 70 years or over; 2) proficient in written and spoken English; and 3) able to provide informed consent at the time of the initial baseline interview. If subjects with cognitive impairment were unable to summarize the study procedures after undergoing the consent process, a care partner also signed the consent form. Exclusion criteria were: 1) inability to participate due to a severe comorbid illness or environmental circumstances; 2) life expectancy less than 12 months; 3) planned nursing home placement or move from the area within 12 months; 4) active substance abuse or dependence; or 5) a severe, uncontrolled mental disorder that would render the individual unable to complete questionnaires. All study procedures were approved by the local Institutional Review Board (IRB). Subjects were recruited between May, 2011 and September, 2013.

Trained research assistants met with participants in their homes or in a clinical or community setting. All subjects were administered the three depression measures (PROMIS-8a, GDS, MADRS). The order of administration of the two self-report measures

of depression were counterbalanced and randomized such that half of the subjects were randomized to complete the PROMIS-8a first and half were randomized to complete the GDS first. The administration of the MADRS followed the completion of the self-report measures for all participants. The GDS-SF was extracted from the GDS. Four participants left items blank on the PROMIS-8a, four had missing data on the MADRS, and eight left items blank on the GDS.

Measures

The PROMIS Depression Scale contains 28 self-report Likert scale items ranging from Never to Always with higher scores indicating more depression (16). The four areas assessed include: 1) negative mood such as sadness, guilt; 2) negative views of the self, including self-criticism, worthlessness; 3) negative social cognition such as loneliness, interpersonal alienation, and 4) decreased positive affect and engagement such as loss of interest, loss of meaning and purpose (17). While depression is also characterized by somatic symptoms such as changes in appetite, sleep, and psychomotor functioning (18), such items were not included in the PROMIS depression item bank due to possible confounds of somatic symptoms when assessing depression in individuals with comorbid physical conditions (19). All 28-items were administered and the 8 items that make up the 8a short form were extracted and used for comparison. Given the high correlation between the short form and the full-item bank ($r = 0.95$, $p < .01$), the short form was utilized in our analyses. The raw scores were converted to T-scores via the conversion table given in the PROMIS scoring manual (17). T-scores on the short form range from 0 to 81.3; 0-55 is “normal”, 55.1-59.9 “mild”, 60-64.25 “moderate”, and over 64.26 “severe” depression. T-scores have a mean of 50 and standard deviation of 10.

The Montgomery-Asberg Depression Rating Scale (MADRS) is a widely utilized 10-item clinician-rated measure for the assessment of depression severity (20). Participants are interviewed and rated on a scale from 0 to 6 on each item with higher scores signifying more severe depression. The MADRS correlates well with another widely utilized depression rating scale, the Hamilton Depression Rating Scale (HAM-D), but is preferred for use with older adults due to its limited reliance on vegetative symptoms such as fatigue or physical complaints (21). The MADRS is frequently utilized in treatment trials and performs well in evaluating change over time (20). Inter-rater reliability with different pairs of raters is in the order of 0.89-0.97. The MADRS has a high internal consistency with a Cronbach's $\alpha = 0.85$. Total MADRS scores range from 0-60; 0-7 normal, 8-15 mild depression, 16-25 moderate, 26-30 severe, and 31+ very severe (20).

The Geriatric Depression Scale (GDS) is a 30-item self-report measure of depression severity in older adults (12). The GDS does not rely on questions which may overlap with the somatic problems that older adults, with or without depression, may experience (22). The scale utilizes yes/no responses, which is easier for older individuals than Likert scales (12) and requires minimal cognitive involvement (23). Total GDS scores range from 0-30; 0-9 normal, 10-19 mild, and 20-30 severe (23). The GDS has demonstrated strong psychometric properties including robust internal consistency (Cronbach's $\alpha = 0.91$), split-half reliability of 0.94, and a test-retest correlation of 0.85 over one month (24). The GDS

has also been used to evaluate depression change over time in older adults (25). The 15 items that make up the GDS Short Form (GDS-SF) were extracted from the 30-item version. There was a high correlation between the GDS full and short form ($r = 0.91$, $p < .01$). GDS-SF scores range from 0-15; 0-4 none, 5-9 mild, and 10-15 moderate to severe depression (12). Analyses were conducted with both the GDS and GDS-SF.

The Saint Louis University Mental Status Exam (SLUMS) is an 11-item screening tool used to evaluate cognitive status in adults. Compared to the Mini-Mental State Exam (MMSE), another screening tool for dementia, the SLUMS has the advantage of identifying individuals with mild cognitive impairment (MCI) (26). The MMSE and the SLUMS have relatively high convergent reliability ($r = 0.75$) in community-dwelling older adults (27). The domains assessed by the SLUMS include orientation, registration/recall, immediate and delayed recall, visuospatial abilities, attention, abstraction, and executive function (28). The SLUMS items can further be divided into three categories: 3 orientations items, 9 reasoning items, and 6 memory items (29). Scores range from 0-30 with higher scores indicating better cognition. Using norms developed by the St. Louis Veteran's Association Medical Center, a score of 25 or higher (27 or higher for high school graduates) falls in the normal range, a score of 20-24 (21-26 for high school graduates) falls in the MCI range, and a score of below 20 (below 21 for high school graduates) falls in the dementia range (30) after taking education into account. The SLUMS has been found to be highly sensitive in both highly- (0.92) and poorly-educated individuals (0.81) (26).

Statistical Analyses

Descriptive statistics described participant characteristics. We categorized the sample into groups according to cognitive status as measured by the SLUMS. Pearson correlations were computed between the PROMIS-8a and Legacy depression measure total scores. To compare the prevalence of depression across measures, we dichotomized and ran frequencies for the four depression measures based on the presence (mild symptoms or above) or absence of depressive symptoms. McNemar tests were run for the PROMIS-8a in combination with each of the other measures in order to assess if it categorized depression differently.

Sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) were calculated between the PROMIS-8a and the MADRS based on presence or absence of mild depression. Additionally, the area under the curve (AUC) of the receiver operating characteristics (ROC) curve was calculated for PROMIS-8a in relation to the MADRS diagnosis to determine its diagnostic capability to match MADRS. We repeated the above analyses comparing the GDS and GDS-SF to the MADRS.

We then conducted item analyses for each combination of measures to determine which specific items best explained any depression classification discrepancies. In each case, we examined those who were categorized as at least mildly depressed on one measure but not mildly depressed on the other. Frequencies were run for responses on each of the test items for only the differentially-classified individuals to identify the items that were endorsed the strongest by the largest percentage of individuals in the discrepant group. After all measures were compared against each other, we then reversed the procedure. For example, when

comparing the PROMIS-8a and GDS, we identified the individuals that were classified as depressed according to the PROMIS-8a yet not according to GDS. We then identified the items on the PROMIS-8a which were endorsed the strongest by the largest percentage of individuals in the discrepant group. We then considered the reverse classification scenario to see how many individuals were classified as not depressed on the PROMIS-8a depression criteria, yet depressed according to GDS criteria and identified the items which best explained the discrepancy.

Mean depression scores were calculated for each of the four measures, and one-way analyses of variance (ANOVAs) were run to determine differences in depression scores by group membership based on cognitive status for the PROMIS-8a, MADRS, GDS, and GDSSF. Additional chi-square analyses were run comparing the three SLUMS categories (normal, MCI, and dementia) by presence or absence of depression for each measure. Finally, internal consistency of each measure was calculated using Cronbach's α to assess if internal validity decreased as cognitive impairment increased. All statistical analyses were conducted with SPSS Version 22.

Results

Sample characteristics are presented in Table 1. In this sample of 304 community-dwelling older adults, there were 61 (20.07%) with normal cognition, 139 (45.72%) with mild cognitive impairment (MCI), and 104 (34.21%) with mild/moderate dementia. The mean age was 78.30 ± 5.74 with the majority being Caucasian (78.62%) and female (71.05%) with at least a high school education (88.82%). The majority of individuals did not have clinically-significant depressive symptoms as measured by the PROMIS-8a T-scores ($M=50.6 \pm 4.6$), MADRS ($M= 4.4 \pm 4.4$), GDS ($M= 5.3 \pm 5.1$), or GDS-SF ($M= 2.0 \pm 2.5$).

Comparison of the PROMIS-8a, MADRS, GDS, and GDS-SF depression measures

The PROMIS-8a, MADRS, GDS, and GDS-SF are all highly correlated (see Table 2). Specifically, the PROMIS-8a was highly correlated with the MADRS ($r=.56$, $df=295$, $p<.01$), GDS ($r=.68$, $df=291$, $p<.01$), and GDS-SF ($r=.60$, $df=291$, $p<.01$).

Categorization of depression based on the PROMIS-8a, MADRS, GDS, and GDS-SF measures

The proportion of individuals with at least mild depression was 62/300 (20.67%) based on the PROMIS-8a, 57/300 (19.00%) based on the MADRS, 53/296 (17.91%) based on the GDS, and 41/296 (13.85%) based on the GDS-SF. Table 3 shows the number of individuals with at least mild depression in each of the three cognitive-status groups as measured by the SLUMS. Exact McNemar tests show there was no difference in the proportion of subjects classified with at least mild depression based on the PROMIS-8a compared to the MADRS (McNemar's $\chi^2(1) = .92$, $p=.41$) or based on the PROMIS-8a compared to the GDS (McNemar's $\chi^2(1) = 1.33$, $p=.31$). However, there was a significant difference in the proportion of subjects classified with at least mild depression as measured by the PROMIS-8a compared to the GDS-SF (McNemar's $\chi^2(1) = 8.00$, $p<.01$) such that the PROMIS-8a identified more depressed individuals than the GDS-SF.

Results pertaining to the sensitivity, specificity, PPV, NPV and the AUC of the ROC curves comparing the PROMIS-8a, GDS, and GDS-SF depression measures to the MADRS-based diagnosis are presented in Table 4.

Table 5 presents the number and percent of individuals dichotomized into the presence or absence of at least mild depression according to the PROMIS-8a and each Legacy depression measure. The results of the item analyses indicate that of those individuals classified as at least mildly depressed on the PROMIS-8a yet not mildly depressed on the MADRS, 13 (48%) reported feeling as if they had nothing to look forward to at least some of the time, and 8 (30%) said they felt hopeless at least some of the time within the previous week on the PROMIS-8a. Of those who were classified as at least mildly depressed on the MADRS but not mildly depressed on the PROMIS-8a, 10 (44%) scored a total of 10 or more and 5 (22%) scored a total of 12 or more on the MADRS. Additionally, 16 (70%) reported having difficulties in starting activities and 11 (48%) reported at least occasional feelings of edginess and ill-defined discomfort according to the MADRS.

For those individuals classified as at least mildly depressed on the PROMIS-8a yet not mildly depressed on the GDS, 14 (50%) reported feeling helpless at least sometimes, 14 (50%) felt as if they had nothing to look forward to at least sometimes, and 12 (43%) reported feeling worthless at least some of the time within the previous week according to the PROMIS-8a. For those individuals classified as at least mildly depressed on the GDS but not mildly depressed on the PROMIS-8a, 12 (60%) said they often felt bored, 12 (60%) preferred to stay home rather than going out and trying new things, 11 (55%) said they have dropped many activities, and 7 (35%) had memory problems on the GDS.

For those individuals classified as at least mildly depressed on the PROMIS-8a yet not mildly depressed on the GDS-SF, the two items that showed the biggest discrepancy were the PROMIS-8a statements “In the past 7 days I felt helpless” and “I felt that I had nothing to look forward to,” which half endorsed as occurring at least sometimes or more frequently. For those individuals classified as at least mildly depressed on the GDS-SF but not mildly depressed on the PROMIS-8a, the biggest item discrepancy was the GDS-SF question asking whether one has a problem with memory, which half endorsed. Additionally, 33% of individuals endorsed the following GDS-SF items: unsatisfied with life, decreased interest in activities, sadness, helplessness and a negative view of one's situation compared to others.

Comparison of MADRS, GDS, GDS-SF, and PROMIS by cognitive sub-groups

Mean depression scores grouped by cognitive functioning ranged from 46.52 ± 6.73 to 48.22 ± 7.90 on the PROMIS-8a, 4.45 ± 5.23 to 4.58 ± 4.86 on the MADRS, 4.38 ± 4.40 to 5.71 ± 5.56 on the GDS, and 1.87 ± 2.30 to 2.23 ± 2.65 on the GDS-SF. There were no statistically significant differences in depression scores by cognitive group on the PROMIS-8a ($F(2,297)=1.20, p=.30$), MADRS ($F(2,295)=1.61, p=.20$), GDS ($F(2,293)=1.30, p=.27$), or GDS-SF ($F(2,293)=.47, p=.63$) measures.

Chi-square analyses comparing PROMIS-8a and each Legacy measure with the three SLUMS categories by absence or presence of at least mild depressive symptoms were all

non-significant (all p values were greater than 0.37). Cronbach's α for each measure ranged from .77 to .90 for each cognitive status group.

Conclusions

With multiple measures to evaluate depression, clinicians assessing older people must decide which measure to use for the population they are testing. Table 6 contrasts some of the features of PROMIS-8a and the Legacy depression measures used in our study. To our knowledge, there have not been any studies either assessing depression using the PROMIS-8a depression scale or comparing it to Legacy measures of depression (the MADRS, GDS, GDSSF) in well-educated community-dwelling older adults with varying levels of cognitive functioning. Previous studies suggest the MADRS is useful regardless of cognitive status (31-33) while the internal consistency of the GDS decreases with increasing severity of dementia (34). Holroyd and Clayton (2000) confirm this limitation by suggesting the GDS “may be adequate for screening in patients with mild cognitive impairment, but not in moderately to severely demented subjects,” (33), p. 3. Similarly, Debruyne et al. (2009) found the GDS was a reliable depression screening tool for patients with MCI, but not those with dementia (35).

The high correlations between measures suggest that all the scales similarly assess the construct of depression in community-dwelling older adults and all have similar predictive validity, yet there are some differences with regard to which individuals are categorized as having at least mild depressive symptoms, depending on the measure used. This observation may be a reflection of the multiple and differing dimensions of the depressive syndrome. The PROMIS-8a identified individuals with at least mild depressive symptoms that the MADRS did not identify based on feeling that they had nothing to look forward to or feelings of hopelessness. It is possible that such symptoms are easier to pick up on in a self-report measure than in an observer-rated scale such as the MADRS. Conversely, the MADRS was more likely to pick up on inner tension and lassitude than the PROMIS.

Those items that differentiated the PROMIS-8a from the GDS and GDS-SF include endorsing feelings of helplessness and not having anything to look forward to. Conversely, the GDS picked up on dropping out of activities, feeling bored, and staying at home. Additionally, both the GDS and GDS-SF identified memory problems. Perceived memory problems may in fact be measuring something other than depression, particularly in those with MCI or dementia.

For those older adults who experience symptoms of depression that are characterized by slowed movements rather than classic sadness or dysphoria (36), the MADRS or GDS might be a better screening measure. Alternatively, since the PROMIS-8a includes a focus on symptoms such as helplessness, worthlessness, and not having anything to look forward to, it might be a better screening measure for depression in more classic dysphoric late-life depression. With regard to which measures are most appropriate for cognitively-impaired individuals, the results suggest that the PROMIS-8a and MADRS both perform well and that the GDS and GDS-SF may be less ideal.

In community-dwelling elders who are mainly non-depressed, as in the current sample, the MADRS appears satisfactory in classifying late-life depression. However, MADRS administration requires a trained rater and more time burden (15-30 minutes typically) than the self-report measures (generally under 10 minutes) and may miss out on feelings of hopelessness or not having anything to look forward to. Cusin et al. (2010) suggest that self-rated measures work better than clinician-rated scales in milder forms of depression since they may be more sensitive in detecting changes (37). The self-reported GDS also seems to perform satisfactorily in identifying individuals who meet the clinical threshold for depression. This could be due to the fact that it is a more in-depth questionnaire or that the short form neglects certain elements that are addressed in the original form. Although the short form takes less time to complete, it may miss some individuals who meet clinically-relevant depression. This conclusion is supported by the finding that the PROMIS-8a identified significantly more individuals as at least mildly depressed than the GDS-SF.

While the question can be raised as to what constitutes clinically-relevant depression, we chose to focus on mild depression as a cutoff rather than moderate depression given that we are evaluating depression in community-dwelling older adults. We wanted to ensure that the depression measure would be sensitive enough to pick up even subtle symptoms of depression. Clinicians can then carry out a more extensive evaluation to determine whether and how to treat the depressive symptoms. This is consistent with the literature which indicates that mild depression is clinically significant (38, 39). Furthermore, Kirchengast and Haslinger indicate that mild depression significantly impacts both the perception of health status and health-related quality of life in older adults (40).

The PROMIS-8a scale takes the least amount of time to complete compared to the MADRS, GDS and GDS-SF measures. Furthermore, the PROMIS-8a seems to identify mild depression as well as the MADRS in older adult samples with varying degrees of cognitive functioning. Our findings suggest that the PROMIS-8a is a sensitive and specific enough tool to use in screening for mild depression. Therefore, given limited resources and high demand, it would be reasonable for clinicians or healthcare systems to consider using the PROMIS-8a as a practical and quick way to assess depression in community-dwelling older adults. How the PROMIS depression scale might perform as a longitudinal assessment, however, is not clear and would need to be further evaluated in populations with more extensive depressive severity that might be expected to fluctuate over time.

Limitations

Notable limitations to the study include the fact that the surveyed population was mainly a euthymic community sample with little ethnic or racial diversity and relatively high levels of education. As such, the results may not be generalizable to groups of older adults who are more depressed, more burdened by medical illness, or more cognitively impaired. Further research should compare the PROMIS and Legacy depression measures in older adult samples with greater demographic and clinical variability. Additionally, the current study only looked at depression cross-sectionally, and thus was unable to evaluate the utility of the PROMIS-8a to identify depression across time.

Acknowledgment

This study was supported by ARRA grant AG038825-01 from the National Institute on Aging (NIA) to Drs. Smyth and Sajatovic. The authors report no financial conflicts of interest for this work.

References

1. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA : the journal of the American Medical Association*. 1989; 262(7):914–9. [PubMed: 2754791]
2. Ownby RL, Crocco E, Acevedo A, et al. Depression and risk for Alzheimer disease: systematic review, meta-analysis, and meta-regression analysis. *Archives of general psychiatry*. 2006; 63(5): 530–8. [PubMed: 16651510]
3. Spira AP, Rebok GW, Stone KL, et al. Depressive symptoms in oldest-old women: risk of mild cognitive impairment and dementia. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*. 2012; 20(12):1006–15. [PubMed: 22015706]
4. Moussavi S, Chatterji S, Verdes E, et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*. 2007; 370(9590):851–8. [PubMed: 17826170]
5. Parashos IA, Stamouli S, Rogakou E, et al. Recognition of depressive symptoms in the elderly: what can help the patient and the doctor. *Depression and anxiety*. 2002; 15(3):111–6. [PubMed: 12001179]
6. Byers AL, Covinsky KE, Barnes DE, et al. Dysthymia and depression increase risk of dementia and mortality among older veterans. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*. 2012; 20(8):664–72. [PubMed: 21597358]
7. Lenze EJ, Schulz R, Martire LM, et al. The course of functional decline in older people with persistently elevated depressive symptoms: longitudinal findings from the Cardiovascular Health Study. *Journal of the American Geriatrics Society*. 2005; 53(4):569–75. [PubMed: 15817000]
8. [2014 16 July] The Patient Reported Outcomes Measurement Information System (PROMIS): A walk through the first four years [Internet]. 2011. Unknown: Unknown; Available from: <http://www.nihpromis.org/>
9. Kjaergaard M, Elisabeth C, Wang A, et al. A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Asberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. *Scandinavian Journal of Psychiatry*. 2014; 55:83–9.
10. Iannuzzo RW, Jaeger J, Goldberg JF, et al. Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. *Psychiatry research*. 2006; 145(1):21–37. [PubMed: 17049379]
11. Bentz BG, Hall JR. Assessment of depression in a geriatric inpatient cohort: A comparison of the BDI and GDS. *Inter J of Clin and Health Psych*. 2008; 8(1):93–104.
12. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*. 1983; 17:37–49. [PubMed: 7183759]
13. Wohlreich MM, Sullivan MD, Mallinckrodt CH, et al. Duloxetine for the treatment of recurrent major depressive disorder in elderly patients: treatment outcomes in patients with comorbid arthritis. *Psychosomatics*. 2009; 50(4):402–12. [PubMed: 19687181]
14. Lelito RH, Palumbo LO, Hanley M. Psychometric evaluation of a brief geriatric depression screen. *Aging & mental health*. 2001; 5(4):387–93. [PubMed: 11767988]
15. Friedman B, Heisel MJ, Delavan RL. Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. *Journal of the American Geriatrics Society*. 2005; 53(9):1570–6. [PubMed: 16137289]
16. Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. *Assessment*. 2011; 18(3):263–83. [PubMed: 21697139]

17. [2014 16 July] PROMIS Scoring Guide [Internet]. 2011. Unknown: Unknown. Available from: www.chiroindex.org/wp-content/uploads/2014/02/PROMIS-Scoring-Guide.pdf
18. Adams KB. Depressive symptoms, depletion, or developmental change? Withdrawal, apathy, and lack of vigor in the Geriatric Depression Scale. *The Gerontologist*. 2001; 41(6):768–77. [PubMed: 11723345]
19. Montorio I, Izal M. The Geriatric Depression Scale: a review of its development and utility. *International psychogeriatrics / IPA*. 1996; 8(1):103–12. [PubMed: 8805091]
20. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *The British journal of psychiatry : the journal of mental science*. 1979; 134:382–9. [PubMed: 444788]
21. Reijnders JS, Lousberg R, Leentjens AF. Assessment of depression in Parkinson's disease: the contribution of somatic symptoms to the clinimetric performance of the Hamilton and Montgomery-Asberg rating scales. *Journal of psychosomatic research*. 2010; 68(6):561–5. [PubMed: 20488273]
22. Jongenelis K, Pot AM, Eisses AM, et al. Diagnostic accuracy of the original 30-item and shortened versions of the Geriatric Depression Scale in nursing home patients. *International journal of geriatric psychiatry*. 2005; 20(11):1067–74. [PubMed: 16250079]
23. Lopez MN, Quan NM, Carvajal PM. A Psychometric Study of the Geriatric Depression Scale. *European Journal of Psychological Assessment*. 2010; 26(1):55–60.
24. Parmalee PA, Lawton MP, Katz IR. Psychometric properties of the Geriatric Depression Scale among the institutionalised aged. *Psychological Assessment*. 1989; 1:331–8.
25. Mackin RS, Insel P, Aisen PS, et al. Longitudinal stability of subsyndromal symptoms of depression in individuals with mild cognitive impairment: relationship to conversion to dementia after 3 years. *Int J Geriatr Psychiatry*. 2012; 27(4):355–63. [PubMed: 21744390]
26. Tariq SH, Tumosa N, Chibnall JT, et al. Comparison of the Saint Louis University mental status examination and the mini-mental state examination for detecting dementia and mild neurocognitive disorder--a pilot study. *Am J Geriatr Psychiatry*. 2006; 14(11):900–10. [PubMed: 17068312]
27. Feliciano L, Horning SM, Klebe KJ, et al. Utility of the SLUMS as a cognitive screening tool among a nonveteran sample of older adults. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry*. 2013; 21(7):623–30. [PubMed: 23567386]
28. Kansagara, D.; Freeman, M. VA Evidence-based Synthesis Program Reports. Washington (DC): 2010. Systematic Evidence Review of the Signs and Symptoms of Dementia and Brief Cognitive Tests Available in VA..
29. Cao L, Hai S, Lin X, et al. Comparison of the Saint Louis University Mental Status Examination, the Mini-Mental State Examination, and the Montreal Cognitive Assessment in detection of cognitive impairment in Chinese elderly from the geriatric department. *Journal of the American Medical Directors Association*. 2012; 13(7):626–9. [PubMed: 22698956]
30. Cruz-Oliver DM, Malmstrom TK, Allen CM, et al. The Veterans Affairs Saint Louis University Mental Status Exam (SLUMS Exam) and the Mini-Mental Status Exam as Predictors of Mortality and Institutionalization. *The journal of nutrition, health & aging*. 2012; 16(7):636–41.
31. Waltis JP, Davies KN, Bunn WK, et al. Correlation between Hospital Anxiety Depression (HAD) scale and other measures of anxiety and depression in geriatric inpatients. *J Psychiatr Res*. 1993; 22:13–9.
32. van Marwijk H, Hoeksema HL, Hermans J, et al. Prevalence of depressive symptoms and depressive disorder in primary care patients over 65 years of age. *Family practice*. 1994; 11(1):80–4. [PubMed: 8034157]
33. Holroyd S, Clayton AH. Measuring Depression in the Elderly: Which Scale Is Best. *MedGenMed*. 2000; 2(4) formerly published in *Medscape Psychiatry & Mental Health eJournal* 5(5), 2000] [Internet].
34. Muller-Thomsen T, Arlt S, Mann U, et al. Detecting depression in Alzheimer's disease: evaluation of four different scales. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*. 2005; 20(2):271–6. [PubMed: 15708735]

35. Debruyne H, Van Buggenhout M, Le Bastard N, et al. Is the geriatric depression scale a reliable screening tool for depressive symptoms in elderly patients with cognitive impairment? *International journal of geriatric psychiatry*. 2009; 24(6):556–62. [PubMed: 19132643]
36. Gallo JJ, Rabins PV. Depression without sadness: alternative presentations of depression in late life. *American family physician*. 1999; 60(3):820–6. [PubMed: 10498109]
37. Cusin, C.; Yang, H.; Yeung, A., et al. *Handbook of Clinical Rating Scales and Assessment*. Humana Press; New York, NY: 2010. Rating scales for depression.; p. 7-37.
38. Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of affective disorders*. 2011; 130(1-2): 26–36. [PubMed: 20708274]
39. Backenstrass M, Frank A, Joest K, et al. A comparative study of nonspecific depressive symptoms and minor depression regarding functional impairment and associated characteristics in primary care. *Comprehensive psychiatry*. 2006; 47(1):35–41. [PubMed: 16324900]
40. Kirchengast S, Haslinger B. Even mild depression reduces health related quality of life (HRQL) among healthy elderly. *J of Med Psych*. 2009; 1(1):3–9.

Table 1**Study Sample Demographics**

Age Mean (SD)	78.3 (5.7)
Female N (%)	216 (71.1)
Race N (%)	
Caucasian	239 (78.6)
African American	57 (18.8)
American Indian/Native Alaskan	1 (0.3)
Asian	3 (1.0)
Native Hawaiian/Pacific Islander	1 (0.3)
Other	3 (1.0)
Ethnicity N (%)	
Hispanic or Latino	3 (1.0)
Not Hispanic or Latino	285 (98.6)
Marital Status N (%)	
Single or never married	24 (7.9)
Married	114 (37.5)
Divorced	57 (18.8)
Widowed	106 (34.9)
Annual Income N (%)	
Less than \$15,000	61 (20.1)
\$15,000-24,999	66 (21.8)
\$25,000-49,999	79 (26.0)
\$50,000+	70 (23.0)
Don't know/refused	27 (8.9)
High School Education or above N (%)	270 (88.8)
Housing Status N (%)	
Private Home	206 (67.8)
Apartment Building	64 (21.1)
Retirement Community (not assisted living)	23 (7.6)
Other	11 (3.6)

^aSLUMS – The Saint Louis University Mental Status Exam

Table 2

Pearson Correlations between the PROMIS-8a and Legacy Measures of Depression with the SLUMS

		PROMIS-8a ^a	MADRS	GDS	GDS-SF
MADRS ^b	Correlation	.56	---		
	Significance (2-tailed)	.000			
	df	295			
GDS ^c	Correlation	.68	.75	---	
	Significance (2-tailed)	.000	.000		
	df	291	290		
GDS-SF ^d	Correlation	.60	.72	.91	---
	Significance (2-tailed)	.000	.000	.000	
	df	291	290	294	
SLUMS ^e	Correlation	-.11	-.06	-.13	-.09
	Significance (2-tailed)	.051	.331	.023	.123
	df	298	298	294	294

^aPROMIS-8a – Patient-Reported Outcomes Measurement Information System Short Form^bMADRS – The Montgomery-Asberg Depression Rating Scale^cGDS – The Geriatric Depression Scale^dGDS-SF - The Geriatric Depression Scale Short Form^eSLUMS – The Saint Louis University Mental Status Exam

Table 3

Number of individuals with at least mild depression as determined by the PROMIS-8a and Legacy depression measures grouped by cognitive status

	PROMIS-8a^a	MADRS^b (N=300)	GDS^c (N=296)	GDS-SF^d (N=296)
SLUMS N (%)				
Normal	9 (3.00)	11 (3.67)	7 (2.36)	8 (2.70)
MCI	25 (8.33)	27 (9.00)	27 (9.12)	18 (6.08)
Dementia	28 (9.33)	19 (6.33)	19 (6.41)	15 (5.07)

^aPROMIS (T-scores): Normal 0-55; Mild 55.1-59.9; Moderate or more 60-81.3

^bMADRS Normal 0-7; Mild 8-15; Moderate or more 16-60

^cGDS: Normal 0-9; Mild 10-19; Moderate or more 20-30

^dGDS-SF: None 0-5; Mild 6-9; Moderate or more 10-15

Table 4

Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) Curve of Self-Report Measures Compared to the MADRS

		PROMIS-8a	GDS	GDS-SF
MADRS	Sensitivity	58.18%	60.00%	50.91%
	Specificity	87.60%	91.56%	94.51%
	PPV ^a	51.61%	62.26%	68.29%
	NPV ^b	90.21%	90.79%	89.24%
	AUC ^c of ROC ^d curve	.79 *	.88 *	.73 *

^a PPV=Positive predictive value

^b NPV=Negative predictive value

^c AUC=Area under the curve

^d ROC= Receiver operating characteristic

* $p < .01$, two-sided Z-test

Table 5

Identification of the Presence or Absence of Mild Depressive Symptoms in the PROMIS-8a versus Legacy Depression Measures

PROMIS-8a		
	Depressed N (%)	Not Depressed N (%)
MADRS N=300		
Depressed	32 (10.67)	23 (7.67)
Not Depressed	27 (9.00)	212 (70.67)
GDS N=293		
Depressed	33 (11.26)	20 (6.83)
Not Depressed	28 (9.56)	212 (72.35)
GDS-SF N=293		
Depressed	26 (8.87)	15 (5.12)
Not Depressed	32 (10.92)	217 (74.06)

Table 6

Comparisons of the Legacy Depression Measures and PROMIS-8a:

	MADRS	GDS	GDS-SF	PROMIS-8a
Question Format	Likert scale	Yes/No format	Yes/No format	Likert scale
Rater	Clinician-rated	Self-rated	Self-rated	Self-rated
Number of items	10 items	30 items	15 items	8 items
Dimensions of depression	Multidimensional	Multidimensional	Multidimensional	Unidimensional
Primary use	Change over time	Screen and change over time	Screen	Screen
Validation In general	Large body of evidence	Large body of evidence	Large body of evidence	Growing body of evidence
In elderly/cognitively impaired	Some evidence	Valid in elderly; not valid in elderly with dementia	Valid in elderly; not valid in elderly with dementia	Growing body of evidence
Training needed	Yes	No	No	No
Burden	High	Minimal	Minimal	Minimal
Additional considerations	Commonly used in clinical trials to measure change	Must meet high severity threshold for affirmative response	Must meet high severity threshold for affirmative response	Accounts for the dysphoric aspect of depression but omits other dimensions like apathy/lethargy