Published in final edited form as:

Int J Geriatr Psychiatry. 2019 May ; 34(5): 700–708. doi:10.1002/gps.5074.

Measurement Validity of the Patient-Health Questionnaire 9 in US Nursing Home Residents

Emmanuelle Belanger, PhD,

Center for Gerontology and Healthcare Research, Department of Health Services, Policy & Practice, Brown University School of Public Health, 121 South Main Street, 6th Floor, Providence, RI, 02903

Kali S. Thomas, PhD,

U.S. Department of Veterans Affairs Medical Center, Providence RI, Center for Gerontology and Healthcare Research, Department of Health Services, Policy & Practice, Brown University School of Public Health

Richard N. Jones, ScD,

Department of Psychiatry and Human Behavior, Brown University Warren Alpert Medical School

Gary Epstein-Lubow, MD, and

Hebrew Senior Life, Harvard Medical School & Department of Psychiatry and Human Behavior Brown University Warren Alpert Medical School

Vincent Mor, PhD

Center for Gerontology and Healthcare Research, Department of Health Services, Policy & Practice, Brown University School of Public Health, U.S. Department of Veterans Affairs Medical Center, Providence RI

Abstract

Objectives: The objective of this study was to assess the measurement properties of the self-reported Patient Health Questionnaire-9 (PHQ-9) and its 10-item observer-version (PHQ-10OV) among nursing home residents.

Methods: We conducted a retrospective study of Minimum Dataset 3.0 assessments for national cohorts of Medicare Fee-for-Service beneficiaries who were newly admitted or incident long-stay residents in 2014-2015 at US Nursing Homes Certified by the Center for Medicare and Medicaid Services. Statistical analyses included examining internal reliability with McDonald's omega,

Corresponding Author: Phone: (401) 863-9924, emmanuelle_belanger@brown.edu.

Publisher's Disclaimer: This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/gps.5074

Conflict of Interest Statement

Vincent Mor is Chair of the Independent Quality Committee at HCR Manor Care, and Chair of the Scientific Advisory Board and consultant at NaviHealth, Inc., as well as former Director of PointRight, Inc., where he holds less than 1% equity. For the remaining authors no conflicts of interest were declared.

Data Sharing and Data Accessibility

CMS data used in this study cannot be shared as per DUA agreements. Analytic code can be accessed at: <https://repository.library.brown.edu/studio/item/bdr:841357/>

structural validity with confirmatory factor analysis, and hypothesis testing for expected gender differences and criterion validity with descriptive statistics. The Chronic Condition Warehouse depression diagnoses were used as an administrative reference standard.

Results: Both the PHQ-9 and PHQ-10OV had good internal reliability with omega values above 0.85. The self-reported scale yielded good model fit for a one-factor solution, while the PHQ-10OV had slightly poorer fit and a lower standardized factor loading on the additional irritability item. Both scales appear sufficiently one-dimensional given that somatic items had higher factor loading on a general depression factor than on a somatic sub-factor. We were unable to obtain expected gender differences on the PHQ-10OV scale. The PHQ-9 and PHQ-10OV were both highly specific but had poor sensitivity compared with an administrative reference standard.

Conclusions: The PHQ-9 appears to be a valid and promising measurement instrument for research about depression among NH residents, while the validity of the PHQ-10OV should be examined further with a structured psychiatric interview as a stronger criterion standard.

Keywords

Depression; Long-Term Care; Psychometrics; Measurement

Introduction

Depression is an important psychiatric disorder in long-term care populations, because of its high prevalence and negative impact on quality of life.^{1,2} The Minimum Data Set 3.0 is a standardized resident assessment in all Medicare and Medicaid certified US nursing homes (NH). It was implemented in 2010 to replace version 2.0, and represents a resource to examine health outcomes and quality indicators among NH residents. The adoption of the Patient Health Questionnaire-9 (PHQ-9) as a standard measure of depression in the MDS 3.0 provides opportunities to examine depressive symptomatology over time.³ While the PHQ-9 is increasingly being used as a measure of depression in NHs, few studies have explored the measurement validity of the scale in this clinical population, particularly when comparing self-reported and observer-versions of this instrument.⁴

The PHQ-9 has good measurement validity in primary care patients where it was originally developed, particularly criterion validity relative to a clinical reference standard and ability to predict loss of functional status and healthcare utilization.^{5,6} A one-dimensional construct is often assumed for the PHQ-9, whereby a total score of 10 and above is considered indicative of depression. A diagnostic algorithm is also used, requiring high scores on both the anhedonia and dysphoria items, and on other symptoms, as well as an indication that depressive symptoms interfere with activities. The validity of the PHQ-9 cannot be automatically assumed for older, institutionalized populations. In fact, the most common PHQ-9 depressive symptoms documented in NH residents are depressed mood, fatigue, sleep problems, and changes in appetite,⁷ many of which represent somatic symptoms that are frequently observed in aging populations with multi-morbidity. The authors of a validation of the PHQ-9 among 713 elderly patients with co-morbidities concluded that the measure had low sensitivity but high specificity when using the algorithm-based scores, and that using a lower cut-off on the total score would improve both sensitivity and specificity.⁸

Moreover, despite the fact that the PHQ-9 is widely used in primary care, its dimensionality has been contested in different clinical populations,^{9–11} which is problematic given the widespread use of the total score as an indicator of depression.

Additional concerns about the validity of the PHQ-9 in NHs emerge from its use among patients suffering from dementia and cognitive impairment, where the Cornell Scale for Depression in Dementia (CSDD) has been used predominantly.¹² In the original validation of the self-reported and observer-version of the PHQ-9 conducted by Saliba and colleagues,⁴ agreement with the modified Schedule for Affective Disorders and Schizophrenia was good for 368 residents without severe cognitive impairment (weighted kappas 0.69). In the small sample of residents with severe cognitive impairment (n= 39 self-reported, n= 48 observer-version), correlations with the CSDD were better than with the previous MDS 2.0 mood items for both the self-reported (0.63 vs. 0.34) and observer-version (0.84 vs. 0.28).⁴ The authors of a study comparing the CSDD and the PHQ-9OV in a sample of 54 US long-term care residents, most of whom were female (87%), further concluded that all participants with mild to severe symptoms on the CSDD were considered to have minimal (total scores 5 to 9) or greater symptoms on the observer version of the PHQ-9, although agreement was higher with a cut-off point at 10 or more.¹³ Overall, validation studies have been conducted in relatively small samples of NH residents with cognitive impairment and limited measurement properties have been examined.

Determining the prevalence of depression and the burden of depressive symptoms is also complicated in NHs, not only because many different instruments have been used to measure depression in geriatric populations, but also because it is important to differentiate between depression that is symptomatic as opposed to recognized or treated depression, all of which may also differ across subpopulations of long-term and post-acute NH residents. According to an international review, the prevalence of depression is estimated to have a wide range; anywhere from 4% to 25% of NH residents may have a major depressive disorder, while the literature suggests that 29% to 82% of residents have minor depression or depressive symptoms.¹⁴ In a study of residents living in 1,492 NHs across five US states in the 1990s, Brown and colleagues found that 11% of residents had an active diagnosis of depression documented in their assessment.¹⁵ In more recent US-based studies, the prevalence of diagnosed depression was 20.3% according to the Medical Expenditure Panel Survey-Nursing Home Component,¹⁶ while it reached 51.8% of long-stay residents with an MDS assessment in 8 states in 2007.¹⁷ The latest study among newly admitted US NH residents in 2011-2013 indicates a prevalence of 26% for an active diagnosis of depression, of which 66-77% had only minimal depressive symptoms on the PHQ-9.⁷ Low depressive symptoms may be explained by underreporting of active diagnoses in the MDS, which is a poorer criterion standard than a structured psychiatric interview, in addition to the high proportion of NH residents receiving antidepressants and who may improve naturally after admission. The adoption of the PHQ-9 in MDS 3.0 represents a promising development to draw conclusions about the prevalence of depression in the NH setting nationally, provided that it proves to have satisfactory measurement validity in this population.

The objective of this study is to assess the measurement properties of the PHQ-9 self-reported and observer-versions among national cohorts of newly admitted and incident long-

term NH residents. Specifically, this investigation focuses on interrelatedness among the PHQ items (internal consistency), the dimensionality of the construct measured (structural validity), the ability to detect expected differences between groups (hypothesis testing), and whether scores reflect a reference standard (criterion validity).¹⁸ We expect that both the self-reported and observer-version will display high internal consistency and a one-dimensional factor structure. We also expect that women will have higher scores than men, and that the established cut-off point of 10 will yield good sensitivity and specificity compared to a reference standard based on the presence of a depression diagnosis in Medicare claims.

Methods

Study Population and Data Sources

Two national cohorts of NH residents age 65 years and above who were Fee-for-Service Medicare beneficiaries were examined: 1) residents newly admitted to a NH in 2014 or 2015 (who did not have a prior nursing home stay in the past), and 2) long-stay residents who became newly long-stay (90 consecutive days without more than 10 days of interruption) during the same study period. For newly admitted residents, we defined the target assessment as the first non-discharge MDS assessment following admission. For the incident long-stay cohort, we defined the target assessment as the first quarterly or change of status assessment conducted after the patient met our long-stay criteria. Only the first admission or long-stay period was considered for any resident with multiple stays. Residents missing the depression score were excluded from this analysis, representing 1.1% and 0.9% of newly admitted and long-stay cohorts, respectively. These percentages include residents with more than two missing items and for whom a total score was not calculated. Given the clinical difficulty of assessing mood in the face of delirium, all residents with a positive Confusion Assessment Method (CAM) score for delirium were also excluded from analyses.¹⁹ We used end-of-year depression flags from the Chronic Condition Warehouse (CCW) corresponding to the year of the target assessment as a reference standard for depression diagnosis. The CCW flags capture the presence of any International Classification of Diseases (ICD) code for depression from a list of 33 ICD-9 codes in 2014 or 32 ICD-10 codes in 2015. The CCW reference period for depression lasts one year and includes claims from inpatient settings, skilled nursing facilities, home health agencies, hospital outpatient payments, and carrier claims. Access to these data was granted under a data use agreement with the Centers for Medicare and Medicaid Services (DUA#18900).

Depressive Symptoms and Other Study Variables

As part of the standard PHQ-9 questionnaire in the MDS, residents are asked to indicate if, over the past two weeks, they have been bothered by any of nine depressive symptoms, including anhedonia, dysthymia, sleeping problems, fatigue, appetite problems, low self-worth, trouble concentrating, motor disturbances, and suicidal thoughts.⁵ If the symptom is present, the resident is then asked to report symptom frequency, which is scored as: 0) never or 1 day, 1) 2-6 days (several days), 2) 7-11 days (half or more of the days), or 3) 12-14 days (nearly every day). The wording of the observer version, which is completed by NH staff, is exactly the same, except for low self-worth and suicidal thoughts which refer to behavioral

manifestations: “indicating that s/he feels bad about self, is a failure, or has let self or family down”, and “states that life isn’t worth living, wishes for death, or attempts to self-harm”. The observer version also contains an additional item about irritability: “Being short-tempered, easily annoyed”, and is henceforth labeled as the PHQ-10OV. The total score for the PHQ-10OV ranges 0 to 30 as opposed to 0 to 27 for the original, self-reported PHQ-9 scale. Other demographic and clinical variables were taken from the same MDS assessment, including age, sex, race and ethnicity, active diagnosis of depression, use of antidepressants, delirium,²⁰ and Cognitive Function Scale.²¹

Statistical Analyses

All analyses were carried out separately for new admissions and incident long-stay residents, and for those with a self-reported PHQ-9 and a PHQ-10OV. We conducted confirmatory factor analysis using weighted least square for ordinal categorical variables using the Mplus (version 8.1, Muthén & Muthén, Los Angeles, CA) software, and compared standardized factor loadings and fit statistics for a one-factor solution and a bi-factor solution that let somatic symptoms (sleep, energy, appetite) load on a sub-factor.²² The one factor model represents a strict test for a one-dimensional model, while the bi-factor model comprises a general depression factor, and a specific somatic sub-factor, thus allowing us to assess whether the general factor explains the majority of variance in individual items that can theoretically be expected to be related.^{23,24} Given large sample sizes that would always favor the saturated model, we assess model fit using the following standard indices: Root Mean Square Error of Approximation (RMSEA), comparative fit index (CFI), and Tucker Lewis Index (TLI). A lower value on the RMSEA indicates a better fit (.05), while larger values (.095) for CFI and TLI indicate good model fit.²² As part of sensitivity analyses, we examined fit statistics across the 2014 and 2015 study samples, respectively, hypothesizing that results should be consistent across years given a time invariant factor structure. We also assessed internal consistency for standardized factor loadings using McDonald’s omega.²⁵

Hypothesis testing was carried out, first with a visual comparison of the distribution of total scores between men and women across the study populations, and then testing for differences in their mean total scores with an ANOVA. We expected to find differences between men and women residing in NHs, as documented in the general population for this age group. According to the US Center for Disease Control and Prevention, in 2009-2012 3.4% of men age 60 and over had depression, compared to 7.1% of women.²⁶ We also verified the specificity, sensitivity, and positive predicted value of a score of 10 or more against the presence of an end-of-year diagnosis of depression in the CCW for the year of target MDS assessment. As part of sensitivity analyses, these statistics were also examined for a total score of 6 and above.⁸ The diagnostic algorithm that is often used in primary care was not used because the additional item about depressive symptoms interfering with function and activities is not available in the MDS 3.0 and is considered less relevant for an institutionalized population with functional impairment.⁴

Results

Description of Study Sample

The sample of new admissions comprised 1,734,785 NH residents from 2014-2015, with 92.9% of them self-reporting their depressive symptoms, and 7.1% having observer-rated depression scores (see Table 1). Among the 429,653 residents in the incident long-stay cohort, 87.3% self-reported their depressive symptoms and 12.7% had a PHQ-10OV. Among newly admitted residents, 25.1% of residents who self-reported their PHQ-9 had an active depression diagnosis in the concurrent MDS assessment record and 25.5% of those with a PHQ-10OV, while respectively 4.6% and 8.3% of them had scores indicative of depression, above or equal 10. The prevalence of active diagnoses and symptomatic depression was much higher in the incident long-stay sample, with 41.8% of the PHQ-9 and 38.2% of the PHQ-10OV samples having an active diagnosis, and 4.4% and 10.3% of them having a total PHQ score above or equal 10. Very few of the residents who self-reported symptoms had severe cognitive impairment (0.3% and 0.6% in new admissions and long-stay). There is a clear shift to the OV version with greater cognitive impairment; among those with a PHQ-10OV assessment, 71.5% of newly admitted residents had moderate to severe cognitive impairment, while this reached 87.3% of long-stay residents. Use of antidepressant medication was reported in over a third of newly admitted residents, and half of all long-stay residents.

Internal Consistency and Structural Validity

Table 2 displays the results of the confirmatory factor analysis for a one-factor solution, with standardized factor loadings and McDonald's omegas. Higher standardized factor loadings are better and indicate that a high proportion of the variance in the individual items is captured with the latent construct. The results indicate loadings that are moderately high overall, with irritability being a notable exception. This item appears on the PHQ-10OV and demonstrated relatively lower factor loadings of 0.45 and 0.48 in the new admission and long-stay cohorts, respectively. As far as the internal consistency of the scales is concerned, good values above 0.86 were obtained on the omega throughout the models. Fit statistics indicate a good model fit for the PHQ-9 with RMSEA values below 0.05, and CFI / TLI values above 0.95. Fit statistics for the PHQ-10OV are also acceptable, except for the RMSEA values of 0.06. Table 3 presents the standardized factor loadings and fit statistics for bi-factor solutions. In all bi-factor models, the standardized factor loadings of somatic items are higher on the general depression factor than on the somatic factor throughout the different study populations, indicating that an underlying depression factor explains the majority of the variance in the somatic items. The bi-factor models improve RMSEA for the PHQ-10OV samples to reach good model fit throughout indicators. These results were stable across the two years included in the sample.

Hypothesis Testing

In terms of prevalence across genders, men had significantly lower rates of diagnosed depression than women in the CCW, in both newly admitted (33.6% vs. 40.7%) and long-stay cohorts (49.4% vs. 54.5%). The mean total scores for the self-reported PHQ-9 were significantly higher among women in both the new admissions (men: 2.39 SD 3.33; women:

2.45 SD 3.28, $F=161.95$, $p<001$), and long-stay residents (men: 2.17 SD 3.22; women: 2.25 SD 3.19; $F=60.21$, $p=0.001$). Contrary to expectations, however, mean total scores were higher for newly admitted male than female residents when they were assessed by an observer (men: 3.03 SD 4.08; women: 2.94 SD 4.01; $F=15.55$, $p<001$), and mean scores did not differ significantly between men and women in the OV long-stay population (men: 3.33 SD 4.30; women: 3.29 SD 4.21; $F=0.88$, $p=.35$). As part of sensitivity analyses, we compared mean OV scores across genders without the irritability item, and although men did score higher on irritability, we did not find expected differences among either the short-stay (men: 2.76 SD 3.85; women: 2.79 SD 3.83; $F=2.28$, $p=0.13$) or the long-stay cohorts (men: 2.99 SD 4.00; women: 3.02 SD 3.96; $F=1.12$, $p=0.29$). Figure 1 further provides a visual comparison of the distribution of PHQ-9 and PHQ-10OV scores for men and women. Men's scores were inversed for ease of comparison within the same histograms. Among both newly admitted and long-stay residents, the distribution of scores for men who self-reported their depressive symptoms is more concentrated toward 0, while the distribution shifts to higher scores among men assessed by NH staff. The distribution among men and women are also visibly more similar in the OV than the self-reported versions.

Criterion Validity

As displayed in Table 4, when compared against the criterion of having a diagnosis of depression during a physician or hospital encounter in the last year, the positive predictive value of a PHQ-9 score of 10 or above was 0.54 and 0.67 for the newly admitted and long-stay populations respectively, and at 0.40 and 0.51 for the PHQ-10OV. Scores above 10 were highly specific (above .90), indicating that there are few false positives on this test. The sensitivity of PHQ scores remained poor across all samples, not exceeding 0.11, suggesting that many of those with a diagnosis of depression in claims data are not symptomatic on the PHQ-9. Sensitivity was only modestly improved by using a lower threshold of 6 on the total score, or by restricting analyses to those without antidepressant use (Table 4).

Discussion

The purpose of this work was to assess the measurement validity for the PHQ-9 as part of MDS 3.0 assessments in US NHs, which represents a different population from the primary care patients where the majority of validation work has been carried out. Our results confirm that the PHQ-9 and PHQ-10OV have good internal consistency, and that a large part of the variance in the individual items of the original scale is captured in a one-dimensional construct, as demonstrated by high standardized factor loadings. The fact that we obtained higher factor loadings than in previous work can be attributed to our use of weighted least squares estimation for ordinal dependent variables, which allows for better modeling of ordinal categorical items than previous work assuming continuous item responses.^{10,27,28} Bi-factor models yielded a slightly better model fit throughout the different study populations, although the somatic items consistently loaded more highly on the general depression factor. The PHQ-9 therefore represents a sufficiently one-dimensional construct, since the preponderance of the variance is indeed attributable to the general depression factor, despite the existence of some multidimensionality for somatic items.²³

Validity appears to be slightly more problematic for the PHQ-10OV because of near acceptable factor loadings for the item about irritability and fit statistics on the one-factor solution, as well as the lack of expected gender differences despite the fact that gender differences in the prevalence of depression have been found to persist in late life.²⁹ One possible reason for slightly poorer model fit could be attributable to the high rates of dementia among residents who have a PHQ-10OV score. A study about the dimensionality of the Cornell Scale for Depression in Dementia (CSDD) concluded that the scale is multidimensional, and points to known complexity in assessing depression in patients with dementia.³⁰ Whether irritability should be considered an intrinsic component of depression presentation among NH residents with dementia remains an important clinical question, given lower factor loadings. Irritability figures among the items measured with the CSDD, and despite criticisms about its complexity and diagnostic accuracy in the NH setting,³¹ irritability emerged among the four items most suitable to assess depression in the validation of an abridged CSDD in NHs.³² Given evidence of slightly poorer measurement validity for the PHQ-10OV and conflicting findings of previous studies,^{4,13,36,37} we recommend larger validation studies comparing the PHQ-10OV with the CSDD against clinical assessment by qualified healthcare providers.

Research using the Geriatric Depression Scale in Australia revealed that among 168 assisted living residents without cognitive impairment, there was an increase in the detection of major depressive disorder from 16% to 22% when considering the responses of care staff³³ It is therefore not surprising to obtain a higher prevalence of major depression when using provider assessments. However, the lack of gender differences casts doubt about the equivalence of self-reported and observer-versions of this scale. Gender and cognitive impairment have both been identified as factors that may reduce healthcare providers' ability to recognize depression.³⁴ NH staff would benefit from additional training on the manifestation of depressive symptoms among individuals with cognitive impairment, especially considering that these measures are incorporated into care planning and quality measurement.

The PHQ-9 and PHQ-10OV both had high specificity but low sensitivity, which is consistent with other studies using diagnoses from administrative claims as a reference standard. Previous work using MDS 2.0 data between 1999 and 2007 had shown that the positive predictive value of an active depression diagnosis in the MDS was poor (0.25) relative to ICD-9 diagnoses in Medicare hospital claims.³⁵ Sensitivity was not improved substantially by using a lower cut-off point of 6 which had been found to be optimal to detect any depressive disorder in a previous study of Dutch older adults,⁸ even though lower cut-points may provide a better depiction of the significant burden of subthreshold depression in this population.³⁸ Previous work carried out among 112 residents and using a clinical assessment by a geriatric psychiatrist as the reference standard suggested that the self-reported PHQ-2 had better specificity and sensitivity than the Geriatric Depression Scale and a modified version of the CSDD for use in long-term-care. Given that the original validation by Saliba and colleagues yielded better sensitivity vis-à-vis other validated scales and structured psychiatric interviews conducted by research nurses,⁴ it would be valuable to replicate their findings in larger samples of NH residents and using structured interviews completed by psychiatrists, clinical psychologists and psychiatric social workers.³⁶

Despite using national populations of NH residents and examining several dimensions of measurement validity, this study has some limitations. The use of secondary administrative data rather than primary clinical assessments entails the use of recognized depression as the reference standard. Moreover, while the MDS indicates how many residents are receiving antidepressant medication, indication for this prescription is not included, and antidepressants are likely used for indications other than depression, such as persistent neuropathic pain. Descriptive statistics confirmed previously documented high rates of antidepressant use in this population,⁷ and a substantial percentage of residents receiving an antidepressant despite not having a diagnosis of depression. We did not systematically exclude residents using antidepressants from our analyses, but these high rates reduce our ability to assess the criterion validity of the PHQ-9.

We sought to provide a comprehensive assessment of measurement validity for the PHQ-9 and PHQ-10OV in the NH setting. Our results indicate good internal reliability for both scales, and acceptable structural validity for a one-factor solution on the self-reported PHQ-9. Despite a sufficiently one-dimensional construct on both scales, whereby somatic items load more highly on a general depression factor, the slightly poorer fit statistics, near acceptable factor loadings on the additional item about irritability, and our inability to obtain expected gender differences raise questions about the validity of the PHQ-10OV. Criterion validity also remains to be demonstrated using a clinical rather than administrative reference standard. We conclude that the PHQ-9 has tremendous potential to advance research about a burdensome psychiatric condition among NH residents, and that the validity of the observer version of the scale should be examined further.

Acknowledgements

This study was supported by a Program Project Grant from the National Institute of Aging (5P01 AG027296-09).

References Cited

1. Dow B, Lin X, Tinney J, Haralambous B, Ames D. Depression in older people living in residential homes. *Int Psychogeriatr*. 2011;23(5):681–99. [PubMed: 21429283]
2. Thakur M, Blazer DG. Depression in long-term care. *J Am Med Dir Assoc*. 2008;9(2):82–7. [PubMed: 18261699]
3. Saliba D, Buchanan J. Making the investment count: revision of the Minimum Data Set for nursing homes, MDS 3.0. *J Am Med Dir Assoc*. 2012;13(7):602–10. [PubMed: 22795345]
4. Saliba D, DiFilippo S, Edelen MO, Kroenke K, Buchanan J, Streim J. Testing the PHQ-9 interview and observational versions (PHQ-9 OV) for MDS 3.0. *J Am Med Dir Assoc*. 2012;13(7):618–25. [PubMed: 22796361]
5. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. Springer-Verlag; 2001;16(9):606–13. [PubMed: 11556941]
6. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345–59. [PubMed: 20633738]
7. Ulbricht CM, Rothschild AJ, Hunnicutt JN, Lapane KL. Depression and cognitive impairment among newly admitted nursing home residents in the USA. *Int J Geriatr Psych*. 2017;32(11):1172–81.
8. Lamers F, Jonkers CCM, Bosma H, Penninx BWJH, Knottnerus JA, van Eijk JTM Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol*. 2008;61(7):679–87. [PubMed: 18538262]

9. Chilcot J, Rayner L, Lee W, Price A, Goodwin L, Monroe B, et al. The factor structure of the PHQ-9 in palliative care. *J Psychosom Res.* 2013;75(1):60–4. [PubMed: 23751240]
10. Forkmann T, Gauggel S, Spangenberg L, Brähler E, Glaesmer H. Dimensional assessment of depressive severity in the elderly general population: Psychometric evaluation of the PHQ-9 using Rasch Analysis. *J Affect Disord.* 2013;148(2-3):323–30. [PubMed: 23411025]
11. Krause JS, Saunders LL, Bombardier C, Kalpakjian C. Confirmatory factor analysis of the Patient Health Questionnaire-9: a study of the participants from the spinal cord injury model systems. *PMR.* 2011;3(6):533–40–quiz540.
12. Snowden M, Sato K, Roy-Byrne P. Assessment and treatment of nursing home residents with depression or behavioral symptoms associated with dementia: a review of the literature. *J Am Geriatr Soc.* 2003;51(9): 1305–17. [PubMed: 12919245]
13. Phillips LJ. Measuring symptoms of depression: comparing the Cornell Scale for Depression in Dementia and the Patient Health Questionnaire-9-Observation Version. *Res Gerontol Nurs.* 2011;5(1):34–42. [PubMed: 22165998]
14. Seitz D, Purandare N, Conn D. Prevalence of psychiatric disorders among older adults in long-term care homes: a systematic review. *Int Psychogeriatr.* 2010;22(7): 1025–39. [PubMed: 20522279]
15. Brown MN, Lapane KL, Luisi AF. The management of depression in older nursing home residents. *J Am Geriatr Soc.* 2002;50(1):69–76. [PubMed: 12028249]
16. Jones RN, Marcantonio ER, Rabinowitz T. Prevalence and correlates of recognized depression in U.S. nursing homes. *J Am Geriatr Soc.* 2003;51(10):1404–9. [PubMed: 14511160]
17. Gaboda D, Lucas J, Siegel M, Kalay E, Crystal S. No longer undertreated? Depression diagnosis and antidepressant therapy in elderly long-stay nursing home residents, 1999 to 2007. *J Am Geriatr Soc.* 2011;59(4):673–80. [PubMed: 21410441]
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737–45. [PubMed: 20494804]
19. Inouye SK, Kosar CM, Tommet D, Schmitt EM, Puelle MR, Saczynski JS, et al. The CAM-S: development and validation of a new scoring system for delirium severity in 2 cohorts. *Ann Intern Med.* 2014;160(8):526–33. [PubMed: 24733193]
20. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among older hospitalized patients. *JAMA.* 1998;279(15): 1187–93. [PubMed: 9555758]
21. Thomas KS, Dosa D, Wysocki A, Mor V. The Minimum Data Set 3.0 Cognitive Function Scale. *Med Care.* 2017;55(9):e68–e72. [PubMed: 25763665]
22. Kline RB. Principles and practice of structural equation modeling, Third Edition Guilford Press; 2010.
23. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient Health Questionnaire Anxiety and Depression Scale. *Psychosom Med.* 2016;78(6):716–27. [PubMed: 27187854]
24. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res.* 2007;16(S1):19–31. [PubMed: 17479357]
25. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol.* 2014;105(3):399–412. [PubMed: 24844115]
26. Pratt LA, Brody DJ. Depression in the U.S. Household Population, 2009–2012. National Center for Health Statistics [Internet]. 2014 11 18 [cited 2018 Aug 28]:[8p] Available from: <https://www.cdc.gov/nchs/data/databriefs/db172.pdf>
27. Elhai JD, Contractor AA, Tamburrino M, Fine TH, Prescott MR, Shirley E, et al. The factor structure of major depression symptoms A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Res.* 2012; 199(3): 169–73. [PubMed: 22698261]
28. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. *J Affect Disord.* 2015; 170(C): 138–42. [PubMed: 25240840]

29. Luppia M, Sikorski C, Luck T, Ehreke L, Konnopka A, Wiese B, Weyerer S, König HH, Riedel-Heller SG. Age- and gender-specific prevalence of depression in latest-life – Systematic review and meta-analysis. *J Affect Disord.* 2012;136:212–221. [PubMed: 21194754]
30. Barca ML, Engedal K, Selbaek G, Knapskog A-B, Laks J, Coutinho E, et al. Confirmatory factor analysis of the Cornell scale for depression in dementia among patient with dementia of various degrees. *J Affect Disord.* 2015;188:173–8. [PubMed: 26363614]
31. Jeon Y-H, Li Z, Low L-F, Chenoweth L, O'Connor D, Beattie E, et al. The clinical utility of the Cornell Scale for Depression in Dementia as a routine assessment in nursing homes. *Am J Geriatr Psychiatry.* 2015;23(8):784–93. [PubMed: 25256214]
32. Jeon Y-H, Liu Z, Li Z, Low L-F, Chenoweth L, O'Connor D, et al. Development and validation of a short version of the Cornell Scale for Depression in Dementia for screening residents in nursing homes. *Am J Geriatr Psychiatry.* 2016;24(11): 1007–16. [PubMed: 27538349]
33. Davison TE, McCabe MP, Mellor D. An examination of the “gold standard” diagnosis of major depression in aged-care settings. *Am J Geriatr Psychiatry.* 2009;17(5):359–67. [PubMed: 19390293]
34. Teresi JA, Abrams R, Holmes D, Ramirez M, Shapiro C, Eimicke JP. Influence of cognitive impairment, illness, gender, and African-American status on psychiatric ratings and staff recognition of depression. *Am J Geriatr Psychiatry.* 2002;10(5):506–14. [PubMed: 12213684]
35. Mor V, Intrator O, Unruh MA, Cai S. Temporal and Geographic variation in the validity and internal consistency of the Nursing Home Resident Assessment Minimum Data Set 2.0. *BMC Health ServRes.* 2011; 11:78.
36. Watson LC, Zimmerman S, Cohen LW, Dominik R. Practical depression screening in residential care/assisted living: five methods compared with gold standard diagnoses. *Am J Geriatr Psychiatry.* 2009;17(7):556–64. [PubMed: 19554670]
37. Teresi J, Abrams R, Holmes D, Ramirez M, Eimicke J. Prevalence of depression and depression recognition in nursing homes. *Soc Psychiatry Psychiatr Epidemiol.* 2001;36:613–620. [PubMed: 11838834]
38. Meeks TW, Vahia IV, Lavretsky H, Kulkarni G, Jeste DV. A tune in “a minor” can “b major”: A review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *J Affect Disord.* 2011;129:126–142. [PubMed: 20926139]

Key points

- The self-reported Patient Health Questionnaire-9 (PHQ-9) and the 10-item observer version (PHQ-10OV) administered as part of the Minimum Dataset have good internal reliability in both newly admitted and long-stay nursing home residents.
- The scales appear to be sufficiently one-dimensional given that somatic items had higher factor loadings on a general depression factor than on a somatic sub-factor, and yielded good model fit statistics for a one-factor solution in confirmatory factor analysis.
- The PHQ-10OV had slightly poorer model fit and a lower standardized factor loading on the additional irritability item, and it did not yield expected gender differences.
- The PHQ-9 and PHQ-10OV were both highly specific but had poor sensitivity compared with an administrative reference standard, and should be examined further against a structured psychiatric interview.

Figure 1:
Distribution of PHQ-9 and PHQ-10OV Scores among Male and Female Nursing Home Residents, Total scores for male residents are reversed for ease of comparison

Table 1:

Descriptive Characteristics of Study Samples in 2014-2015, n (%)

	New Admissions		Incident Long-Stay Cohort	
	N= 1,734,785		N= 429,653	
	PHQ-9	PHQ-10OV	PHQ-9	PHQ-10OV
	n= 1,611,836	n= 122,949	n= 374,947	n= 54,706
Age				
65-74	447,662 (27.77)	26,323 (21.41)	75,458 (20.12)	9,740 (17.80)
75-84	589,572 (36.58)	43,440 (35.33)	120,314 (32.09)	18,620 (34.04)
85+	574,602 (35.65)	53,186 (43.26)	179,175 (47.79)	26,346 (48.16)
Gender				
Men	617,761 (38.33)	51,016 (41.49)	131,250 (35.00)	19,126 (34.96)
Women	994,075 (61.67)	71,933 (58.51)	243,697 (65.00)	35,580 (65.04)
Race/Ethnicity				
White	1,341,765 (83.24)	92,195 (74.99)	310,090 (82.70)	41,089 (75.11)
Black	120,902 (7.50)	12,926 (10.51)	33,252 (8.87)	6,276 (11.47)
Hispanic	51,216 (3.18)	6,670 (5.43)	13,264 (3.54)	3,095 (5.66)
Asian	26,023 (1.61)	5,071 (4.12)	5,571 (1.49)	2,236 (4.09)
Other / Unknown	71,930 (4.46)	6,087 (4.95)	12,770 (3.40)	2,010 (3.67)
Cognitive Function				
Cognitively intact	1,049,209 (65.09)	5,735 (4.66)	154,048 (41.09)	836 (1.53)
Mildly impaired	340,053 (21.10)	25,455 (20.70)	108,052 (28.82)	5,191 (9.49)
Moderately impaired	216,539 (13.43)	47,887 (38.95)	110,317 (29.42)	25,014 (45.72)
Severely impaired	4,129 (0.26)	39,969 (32.51)	2,079 (0.55)	22,739 (41.57)
MDS Depression				
Active Diagnosis	404,369 (25.09)	31,326 (25.48)	156,620 (41.77)	20,915 (38.23)
Antidepressants	506,623 (31.43)	40,639 (33.05)	197,061 (52.56)	27,334 (49.97)
PHQ Score 10	73,860 (4.58)	10,138 (8.25)	16,465 (4.39)	5,627 (10.29)

PHQ: Patient Health Questionnaire, OV: Observer Version

Table 2:

Standardized Factor Loadings and Fit Statistics for a One-Factor Solution

PHQ Items	New Admissions		Incident Long-Stay Cohort	
	PHQ-9	PHQ-10OV	PHQ-9	PHQ-10OV
Anhedonia	0.67	0.77	0.67	0.81
Dysphoria	0.71	0.66	0.72	0.73
Sleeping.	0.61	0.72	0.63	0.79
Energy.	0.74	0.79	0.73	0.83
Appetite	0.61	0.69	0.58	0.68
Self-worth	0.69	0.54	0.70	0.66
Concentration	0.61	0.65	0.58	0.63
Motor	0.58	0.61	0.57	0.61
Suicidality	0.67	0.54	0.63	0.75
Irritability		0.45		0.48
Fit Statistics				
Chi2	76,151	13,782	16,671	5,428
Root Mean Square	0.04	0.06	0.04	0.05
Error of Approximation				
Comparative Fit Index	0.97	0.96	0.97	0.96
Tucker-Lewis Index	0.96	0.95	0.96	0.95
McDonald's Omega	0.87	0.88	0.87	0.87

PHQ: Patient Health Questionnaire, OV: Observer Version

Reference values for good fit statistics: RMSEA .05, CFI .95, TLI .95, Omega .90

Table 3:

Standardized Factor Loadings and Fit Statistics for a Bi-Factor Solution

PHQ Items	New Admissions		Incident Long-Stay Cohort	
	PHQ-9	PHQ-10OV	PHQ-9	PHQ-10OV
Anhedonia	0.70	0.81	0.69	0.81
Dysphoria	0.75	0.69	0.76	0.71
Sleeping - depression	0.52	0.62	0.55	0.63
<i>Sleeping - somatic</i>	<i>0.37</i>	<i>0.37</i>	<i>0.35</i>	<i>0.38</i>
Energy - depression	0.65	0.68	0.65	0.70
<i>Energy - somatic</i>	<i>0.47</i>	<i>0.59</i>	<i>0.49</i>	<i>0.62</i>
Appetite - depression	0.51	0.59	0.50	0.55
<i>Appetite - somatic</i>	<i>0.39</i>	<i>0.35</i>	<i>0.36</i>	<i>0.31</i>
Self-worth	0.72	0.56	0.72	0.57
Concentration	0.63	0.69	0.60	0.64
Motor	0.60	0.64	0.59	0.58
Suicidality	0.69	0.56	0.64	0.61
Irritability		0.47		0.46
Fit Statistics				
Chi2	26,883	8,453	7,710	3,317
Root Mean Square Error of Approximation	0.03	0.05	0.03	0.04
Comparative Fit Index	0.99	0.97	0.99	0.97
Tucker-Lewis Index	0.99	0.96	0.98	0.96

Loadings on the secondary somatic factor are show in italic typeface

PHQ: Patient Health Questionnaire, OV: Observer Version

Reference values for good fit statistics: RMSEA .05, CFI .95, TLI .95, Omega .90

Table 4:

Sensitivity, Specificity and Positive Predictive Value of PHQ-9 and PHQ-10OV against CCW Depression Diagnosis* in Full Cohorts and Sub-Sample without Antidepressants

	PHQ Total Score 10			PHQ Total Score 6		
	Sensitivity	Specificity	PPV	Sensitivity	Specificity	PPV
New Admissions						
PHQ-9	0.07	0.97	0.54	0.19	0.88	0.48
PHQ-10OV	0.09	0.92	0.40	0.22	0.80	0.40
Long Stay						
PHQ-9	0.05	0.97	0.67	0.16	0.89	0.63
PHQ-10OV	0.11	0.91	0.51	0.26	0.77	0.50
New Admissions-No Antidepressants						
PHQ-9	0.06	0.97	0.34	0.18	0.88	0.29
PHQ-10OV	0.08	0.92	0.25	0.22	0.81	0.24
Long Stay- No Antidepressants						
PHQ-9	0.05	0.97	0.41	0.14	0.90	0.37
PHQ-10OV	0.11	0.92	0.31	0.25	0.79	0.29

PHQ: Patient Health Questionnaire, OV: Observer Version, PPV: Positive Predictive Value

* Any ICD-9 depression diagnoses over the past year: 296.20-6, 296.30-6, 296.50-6, 296.60-6, 296.89, 298.0, 300.4, 309.1, 311