Published in final edited form as:

J Clin Nurs. 2017 December ; 26(23-24): 4915–4926. doi:10.1111/jocn.13974.

Depression recognition and capacity for self-report among ethnically diverse nursing homes residents: Evidence of disparities in screening

Audrey Chun, MD^{1,2} [Associate Professor], Joann P. Reinhardt, PhD^{1,2} [Director of Research, Associate Clinical Professor], Mildred Ramirez, PhD^{3,4} [Associate Director], Julie M. Ellis, RN, RM, PhD⁵ [Senior Lecturer], Stephanie Silver, MPH^{3,4} [Project Coordinator], Orah Burack, MA¹ [Senior Research Associate], Joseph P. Eimicke, MS^{3,4} [Data Manager], Verena Cimarolli, PhD¹ [Senior Research Scientist], and Jeanne A. Teresi, EdD, PhD^{3,4,6} [Administrator and Director, Senior Research Scientist]

¹Research Institute in Aging, The New Jewish Home, New York, NY, USA

²Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

³Research Division, Hebrew Home at Riverdale, RiverSpring Health, Riverdale, NY, USA

⁴Division of Geriatrics and Palliative Medicine, Weill Cornell Medical College, New York, NY, USA

⁵School of Nursing and Midwifery, La Trobe University, Melbourne, Vic., Australia

⁶Columbia University Stroud Center, New York State Psychiatric Institute, New York, NY, USA

Abstract

Aims and objectives—To examine agreement between Minimum Data Set clinician ratings and researcher assessments of depression among ethnically diverse nursing home residents using the 9-item Patient Health Questionnaire.

Background—Although depression is common among nursing homes residents, its recognition remains a challenge.

Design—Observational baseline data from a longitudinal intervention study.

Correspondence: Mildred Ramirez, Research Division, Hebrew Home at Riverdale, RiverSpring Health, Riverdale, NY, USA. milramirez@aol.com.

ORCID

Mildred Ramirez <http://orcid.org/0000-0003-1032-1273>

Julie M. Ellis <http://orcid.org/0000-0002-8140-1150>

CONFLICT OF INTEREST

None.

CONTRIBUTIONS

Prepared the first draft of the manuscript and conducted the in-depth interviews: AC; supervised the preparation of the manuscript and provided editorial comments: JR; contributed to the discussion of the results related to the subgroup analyses of Spanish speakers: MR; contributed to the writing of the manuscript and edited the final draft of the manuscript: JE; provided the final editorial comments and supervised the data collection: SS; performed data collection: OB; performed the data analyses and prepared the tables in the manuscript: JPE; reviewed the manuscript: VC and provided overall supervision of the analyses, wrote sections of the manuscript and edited the final draft of the manuscript: JAT. All have reviewed and edited the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Methods—Sample of 155 residents from 12 long-term care units in one US facility; 50 were interviewed in Spanish. Convergence between clinician and researcher ratings was examined for (i) self-report capacity, (ii) suicidal ideation, (iii) at least moderate depression, (iv) Patient Health Questionnaire severity scores. Experiences by clinical raters using the depression assessment were analysed. The intraclass correlation coefficient was used to examine concordance and Cohen's kappa to examine agreement between clinicians and researchers.

Results—Moderate agreement ($\kappa = 0.52$) was observed in determination of capacity and poor to fair agreement in reporting suicidal ideation ($\kappa = 0.10$ – 0.37) across time intervals. Poor agreement was observed in classification of at least moderate depression ($\kappa = -0.02$ to 0.24), lower than the maximum kappa obtainable (0.58 – 0.85). Eight assessors indicated problems assessing Spanish-speaking residents. Among Spanish speakers, researchers identified 16% with Patient Health Questionnaire scores of 10 or greater, and 14% with thoughts of self-harm whilst clinicians identified 6% and 0%, respectively.

Conclusion—This study advances the field of depression recognition in long-term care by identification of possible challenges in assessing Spanish speakers.

Relevance to clinical practice—Use of the Patient Health Questionnaire requires further investigation, particularly among non-English speakers. Depression screening for ethnically diverse nursing home residents is required, as underreporting of depression and suicidal ideation among Spanish speakers may result in lack of depression recognition and referral for evaluation and treatment. Training in depression recognition is imperative to improve the recognition, evaluation and treatment of depression in older people living in nursing homes.

Keywords

depression; disparities; ethnicity; long-term care; Minimum Data Set; PHQ-9; self-report capacity

1 | INTRODUCTION

Depression is one of the most common mental health issues for older adults, particularly among those residing in nursing homes. According to the U.S. Centers for Disease Control (CDC), between 2009 and 2012, about 7.6% of community-dwelling Americans aged 12 and older were estimated to have moderate to severe depressive symptoms. Of these, 3% evidenced severe depressive symptoms, of whom about one-third actually received treatment (Pratt & Brody, 2014). Among community-dwelling Hispanics in the U.S.A., the estimates of depressive symptomatology range from 26% (Pratt & Brody, 2014) to 47% (Chung et al., 2003). The prevalence of depression is even higher in nursing homes and other long-term care settings where rates of major depression can be anywhere from 6% to 24% as estimated by multiple international studies (Bell & Goss, 2001; Brown, Lapane, & Luisi, 2002; Cohen, Hyland, & Kimhy, 2003; Jongenelis et al., 2009; Kramer, Allgaier, Fejtikova, Mergl, & Hegerl, 2009; Parmelee, Katz, & Lawton, 1989; Samuels & Katz, 1995; Teresi, Abrams, Holmes, Ramirez, & Eimicke, 2001). Of note, studies in various nursing homes in London and in the USA have found very high rates of significant depressive symptoms ranging from 38% to 44% (Brown et al., 2002; Cohen et al., 2003; Jongenelis et al., 2009; Kerber, Dyck,

Culp, & Buckwalter, 2005; Kramer et al., 2009; Mann, Graham, & Ashby, 1984; Parmelee et al., 1989; Teresi et al., 2001).

2 | BACKGROUND

Until the late 1980s, many long-term care facilities such as nursing homes lacked standardised ways of assessing depression among their residents and often struggled to initiate quality improvement strategies (Sales et al., 2012). Even in a more recent study of palliative care (McCabe, Mellor, Davison, Hallford, & Goldhammer, 2012), one-third of the participating nurses reported low confidence in recognising depression, knowing the signs and symptoms, and understanding its effects. Although standardised assessment measures have been developed, insufficient staff training remains a problem and there persists the need for research on how practitioners can more effectively use these assessments and the resulting data for referral and treatment.

A standardised depression screening tool, the 9-item Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001), was introduced as a component of the MINIMUM DATA SET, version 3.0 (MDS 3.0) in October 2010 (Saliba & Buchanan, 2008). The MDS (Morris et al., 1990) is a set of internationally used (Chen et al., 2016; Tarride et al., 2016) instruments to conduct a comprehensive assessment of nursing home residents, and its use is federally mandated in the USA for all certified institutions receiving Medicaid and Medicare funding. Medicaid is a joint federal and state programme to assist persons with limited resources with medical costs; Medicare is a federal health insurance programme for people aged 65 and older, and for those with disabilities or end-stage renal disease. The MDS informs treatment planning and provides a payment mechanism and data used to monitor system-wide quality. Assessments are performed on a regular basis including annually, quarterly and if there is a significant change (e.g., decline in two areas such as weight loss and nutritional status). MDS assessment results are used to develop, review and revise residents' care plans as needed. MDS data are gathered by interdisciplinary team members, with data collection and management often supervised by the Resident Assessment Utilization Management department.

The PHQ-9 is based on the diagnostic criteria for major depressive disorder (MDD) in the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV; Simon & Von Korff, 2006). The main purpose of the PHQ-9 is to identify symptomatology during the past 2 weeks that may be indicative of depression and that may trigger actions such as referrals for further assessment. Additionally, the severity score can be used for classification and identification of appropriate therapeutic interventions. Finally, the suicide item, the most severe consequence of depression, is important in the nursing home setting because older adults, particularly males over age 85, have the highest completed suicide rate of any age group (CDC, 2010; Delaney et al., 2011; Kramer et al., 2009).

Standardised assessments such as the PHQ-9 provide information for criteria-based classifications of depression that are less dependent on the subjective judgement of clinicians. Saliba et al. (2012) evaluated the feasibility and validity of the PHQ-9, as well as the PHQ-9 OV (observational version for patients requiring a staff proxy) as screening

measures for mood disorders in nursing home populations (Kroenke et al., 2001). They found that compared with the MDS 2.0 observational items, the PHQ-9 interview (administered by research nurses) had greater agreement with standard criterion measures for depression and the same result was observed for the PHQ-9 OV. Saliba et al. (2012) also reported that 86% of the nurses who participated in the study later expressed in a survey that the PHQ-9 provided new insight into the residents' mood. This is consistent with prior research that found significantly lower rates of depression recognition by nursing home clinicians compared with their psychiatrist counterparts, with only 37%–45% of cases of depression diagnosed by psychiatrists recognised by the clinicians (Teresi et al., 2001). In that study, use of a formal assessment measure (Teresi, Abrams, & Holmes, 2000) enhanced recognition.

Ideally, standardisation should result in greater consistency between MDS and research ratings. However, because the PHQ-9 is embedded within the MDS, which is also used for reimbursement decisions as well as to trigger additional assessments, there is the potential for compromised validity of assessment.

Minimum Data Set data collectors may vary in their training and interviewing and data coding skills and in attitudes towards such assessments. This may in turn result in differences in the ways in which they rate a resident's depression even while using the same standardised instrument. Therefore, there may be variation in the levels to which nursing home staff members implemented the new MDS guidelines (Simons et al., 2012). Connolly, Downes, Reuter, and Fogler (2010) and Simons et al. (2012) found that nursing home social workers expressed discomfort with PHQ-9 items particularly that related to suicidal ideation. These authors also reported large caseloads as a barrier to adequate psychosocial assessments and discussed the need for stronger governmental regulations requiring consistent training of clinicians delivering the PHQ-9 and adequate staffing in order to effectively meet residents' psychosocial needs as revealed by the PHQ-9.

Rationale for the study: The purpose of this study was to examine the performance of the measure in an ethnically diverse sample. MDS clinician raters regularly assess depression using the PHQ-9 as part of quarterly MDS 3.0 assessments; the aim of this study was to compare these ratings to those obtained using the PHQ-9 in a research protocol delivered by trained researchers. The MDS clinician ratings and those obtained by the research staff were also compared with respect to how often they determined that a resident was capable of a self-report interview. A direct resident interview is preferable because it could improve the detection of a possible mood disorder. However, clinicians use the staff interview when a given resident is not capable of understanding or communicating their responses to the questionnaire. Because clinicians determine whether or not a resident is capable of participating in a self-report interview, the consistency of their clinical judgements with those of researchers making this same decision using a standardised capacity screen was also assessed.

3 | METHODS

A mixed method approach (Creswell, 2003) was used in order to obtain a more comprehensive understanding of the issues related to the congruency of MDS clinician and researcher ratings of residents' depression. The quantitative and in-depth interviews were conducted concurrently. The qualitative information obtained complimented the quantitative data and provided deeper insight, and contextual interpretation of the quantitative results.

3.1 | Aims

This study examined the agreement between MDS clinician and researcher ratings of depression in terms of the following: (i) determination of resident capacity for self-report; (ii) report of suicidal ideation; (iii) classification of at least moderate depression (score of 10+); and (iv) total PHQ-9 severity score. In-depth interviews were also conducted to inform the quantitative analyses and findings. For the in-depth interview component of this study, clinician experiences with the PHQ-9 were solicited in terms of its utility as an assessment measure for identifying depressive symptoms in nursing home residents. Because the interview data identified potential problems in assessing the Spanish speakers, a subgroup analysis of Spanish speakers was performed. To our knowledge, no literature exists on the use of the PHQ-9 among Spanish speakers in nursing homes.

This study was approved by the Institutional Review Boards of both participating institutions, protocol numbers 2011-04 and 03/01/P074. A waiver was obtained to collect basic demographic data reported in the aggregate on the subsample of those who were unable to provide informed consent. All protected health information was excluded from the file containing this information. These data were used only for comparison with those who were able to provide informed consent.

3.2 | Sample selection and inclusion/exclusion criteria

The sample for this study was from a clinical trial of a depression recognition training programme for front-line long-term care staff targeting residents from 12 long-term care units in one urban long-term care facility. Baseline observational data collected prior to randomisation were used for these analyses. Residents receiving hospice care were excluded from the sample. The research protocol was to exclude from the study those who were not capable of providing informed consent and thus not able to self-report depression. The capacity screen excludes respondents with severe cognitive impairment.

3.2.1 | Response rate—A total of 227 residents were screened for inclusion. It was determined that 72 residents were unable to consent (27%). An additional 30 residents (12%) refused participation; 155 (59%) were able to self-report depression through direct assessment and consented to participate. Five respondents were missing research PHQ-9 data; thus, some of the analyses were conducted using a reduced sample.

3.3 | Measures

3.3.1 | Capacity to self-report and cognition level—Minimum Data Set clinician assessors determined capacity to self-report based on their clinical judgement, whereas

researchers used the capacity assessment embedded in the Institutional Comprehensive Assessment and Referral Evaluation (INCARE; Golden, Teresi, & Gurland, 1984; Gurland & Wilder, 1984; Teresi, Golden, & Gurland, 1984) administered prior to delivering the PHQ-9.

3.3.2 | MDS capacity guidelines—As stated in the MDS 3.0 Training Manual, Appendix D, “All residents capable of any communication should be asked to provide information.” In addition, it states, “A simple performance-based assessment of cognitive function can quickly clarify a resident’s cognitive status.” Further details about the cognitive scale or cut-off score to indicate level at which administration should be performed are not specified. Per the July 2010 directions, if the resident is rarely or never understood, a staff assessment should be performed. In practice, staff uses clinical judgement in determining whether an individual is “comatose” or “unable to understand,” in which case no further direct assessment is conducted.

3.3.3 | Research capacity screen—The six items to determine capacity to self-report and provide informed consent assess orientation to place and memory. They are as follows: “How old are you?”; “In what year were you born?”; “And the month (when you were born)?”; “Where are we now?”; “What do people do here?”; “What is the name of this place?”; or “What state are we in?”. The latter is offered as an alternative to the previous item for community-residing respondents. The items are scored in the impaired (incapable) direction and summed for a total score. One point is attributed to each incorrect answer; the theoretical score range is 0–6 with six indicative of inability. Respondents who score between 0 and 3 are eligible for completing the full study protocol including extended self-report measures. Those who score 4 or 5 are eligible for only cognitive screening measures, performance of simple ADL actions, and for informant measures, chart reviews and observations. Respondents with a score of 6 are deemed unable to give consent; family or proxy consent is, thus, required. The measure has been used in over 20 studies of approximately 6,500 residents of long-term care facilities, including assisted living, to determine capacity to self-report and provide informed consent for minimally invasive research. The explained common variance (ECV) estimates from factor analyses ranged from 44.22 to 59.86 across studies and was estimated at 42.24 for this study. The McDonald’s omega total reliability estimates ranged from 0.88 to 0.96 across studies and was estimated at 0.92 for this sample, providing evidence for unidimensionality and high reliability. All residents who were capable of consenting based on these criteria were administered the PHQ-9.

3.3.4 | Research cognition assessment—Cognition was assessed to provide a profile of cognitive function in this sample, distinct from capacity to self-report and provide informed consent. The Care Dementia Diagnostic Scale (CAREDiag) is a measure that performs well among ethnically diverse groups (Teresi, Kleinman, et al., 2000). The scale results in several classifications: none (0–3), mild (4–6), moderate (7–10), severe (11–15) and very severe (not testable). The scale was used as the screening measure in two large epidemiological studies of dementia among Hispanics, African Americans and non-Hispanic Whites in North Manhattan. Cronbach’s alphas were 0.83, 0.84 and 0.83, for the Hispanic,

African American and White subgroups, respectively (Gurland et al., 1997). The scale was developed to have maximal relationship with diagnosis and to be culture fair (Golden, Teresi, & Gurland, 1983). The sensitivity estimate of the scale with clinical diagnosis for one of the studies was 0.87, and the specificity was 0.79 (Wilder et al., 1995). The scale has performed well in previous analyses (as compared with six other screening scales) in terms of area under the curves (AUCs) among Hispanic, African American, and non-Hispanic White subgroups and education subgroups, and is relatively unbiased across education and race/ethnic groups (Teresi, Kleinman, et al., 2000). For this sample, the CAREDiag items evidenced Cronbach's and ordinal alphas of 0.77, and 0.84, respectively. The McDonald's omega total reliability estimate was 0.86, and the ECV was 26.86%.

3.3.5 | MDS PHQ-9—The PHQ-9 is Section D of the quarterly MDS 3.0 assessment and is used to assess resident mood. The PHQ-9 contains nine items from the DSM-IV diagnostic criteria for depression shown in the Appendix S1. An attempt is made to first interview residents directly and obtain their responses by self-report, but if that is not possible due to poor cognitive status, the MDS clinician raters interview an informant (usually a certified nurse assistant who knows the resident best) using the Staff Assessment of Resident Mood (PHQ-9 OV). For residents capable of a self-report interview, MDS clinician raters administer the PHQ-9 directly to the resident. They assess both the mood state using the PHQ-9 items and cognitive status, but do not necessarily ask other items at that time.

3.3.6 | Research PHQ-9—The research staff used the standardised research version of the PHQ-9. While the items are ostensibly the same for the two administrations, locally collected MDS data may vary across assessors, if a standardised format of response coding is not used consistently.

3.3.7 | Evaluation of depression assessment by clinicians—Interview items were developed based on focus groups with MDS clinician raters. These items were reviewed by a bilingual psychologist who is expert in qualitative methods. A research fellow in geriatrics administered the items independently in a semi-structured format.

3.4 | Procedures

Residents who consented to participate received in-person interviews by independent research staff. Each of the 155 residents was interviewed in-person with a depression assessment that included the PHQ-9. Research assessors were introduced to the resident by a staff person, after determining the best time to approach the resident.

Minimum Data Set data were obtained from the electronic medical record. The MDS PHQ-9 data were collected as part of the usual assessment by MDS clinician raters.

Fifty of the direct resident research interviews were conducted by Spanish-speaking research interviewers using the standardised Spanish translation of the PHQ-9 available from Pfizer. For the MDS, Spanish-speaking residents were assessed by the MDS clinician raters using an interpreter who translated items from the English version of the PHQ-9.

The research assessments were conducted at the baseline time point prior to the training intervention, and the MDS assessment completed closest to this date in the resident's medical records was used for comparison. The amount of time between these two assessments was recorded and used as a control variable.

Interview data of a qualitative nature were collected concurrently and independently from ten MDS clinician raters to obtain an in-depth perspective on using the PHQ-9 for assessing depression in nursing home residents. They were interviewed one-on-one for 30-45 minutes using 10 open-ended questions.

3.5 | Analysis

The analyses were based only on those 155 residents enrolled in the study; however, demographic data collected as part of the chart review under the IRB waiver were reported in the aggregate. Additionally, MDS-research protocols were assessed in terms of convergence of decisions related to capacity for direct assessment.

3.5.1 | Quantitative—The convergence between the MDS clinician raters and the research assessment was examined for both severity ratings and for depression classification. Additionally, the item indicative of suicidal ideation was examined separately because residents answering affirmatively to the suicidal ideation item are referred for further assessment. A cut-off score of 10 on the PHQ-9 was used for depression classification because this serves as an indicator of moderate depression and triggers further assessment of the resident using the MDS.

As in other validation studies of the MDS using research nurses as the gold standard assessor (e.g., Morris et al., 1990), the research administration was considered the gold standard method for obtaining the PHQ-9 data.

The intraclass correlation coefficient (ICC) was used to examine the association between the total PHQ-9 scores obtained by MDS clinician raters and researchers. Cohen's (Cohen, 1960) kappa was used to assess agreement between MDS clinician raters and researchers in terms of the following: (i) determination of resident capacity for self-report (y/n); (ii) report of suicidal ideation (y/n); and (iii) classification of at least moderate depression (score of 10+; y/n). Additionally, sensitivity, specificity, positive and negative predictive values (ppv, npv) as well as maximum kappa were reported. Although guidelines for interpretation of kappa vary, we used the guidelines of Landis and Koch (1977): <0 no agreement; 0–0.20 slight/poor; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1 almost perfect, recognising that these coefficients are affected by the base rate of the phenomenon examined. However, comparisons with the maximum kappa obtainable, given the marginal distributions, can inform interpretation.

Sensitivity analyses were conducted examining the time interval between research and MDS assessments and mode of assessment (direct only). Finally, subgroup analyses were performed examining agreement among the Spanish speakers.

3.5.2 | Interviews with Clinicians—For the in-depth interviews, responses to the open-ended questions were transcribed before the coding process. The process of developing a coding system for the responses progressed in several stages. A basic set of definitions and rules were applied to the first set of five interview transcripts. Coding decisions were compared with those of a senior research scientist to discuss disagreements as well as to refine previous definitions and rules. The refined set of definitions and rules was used to code the remaining set of five interview transcripts and review the codes already assigned to the first set of five interview transcripts. The rate of agreement was 89%, and any disagreements in coding assignments were resolved by discussion. The frequencies of the key responses given by the MDS clinician raters were then tabulated.

4 | RESULTS

4.1 | Sample characteristics

The sample was comprised of 155 residents; the majority (63.2%) was female. The residents ranged in age from 68 to 91 years (mean = 79.93, $SD = 11.08$); 41.3% were Black (non-Hispanic), and 58 (37.4%) were Hispanic, of whom 50 were interviewed in Spanish. The majority was not married: 23.9% had never been married, and 42.6% were widowed.

Assessed directly by research staff, individuals with severe cognitive impairment and communication disorder were excluded from the sample. Almost one-third (30.3%) of the sample assessed by the research staff were classified as unimpaired, scoring 0–3 on the cognitive screen; 37.4% were classified as mild (scoring 4–6); and 29.7% were classified as moderately cognitively impaired (scoring 7–10). Only one person was classified with severe impairment. Thus about two-thirds (69%) were classified as none or mildly impaired. The mean score was 4.90 ($SD = 2.59$), indicative of mild impairment.

Of the 72 residents who were assessed as without capacity to provide informed consent, all were severely cognitively impaired and did not meet the research criterion for capacity for direct assessment by the research assessors. Among this group, 65% were female and the average age was older than those assessed (83.83, $SD = 10.67$). About half (45.8%) were Black (non-Hispanic) and 25.0% Hispanic (data not shown).

4.2 | Evaluation of depression assessment by clinicians

The 10 in-depth interviews with MDS clinician raters who regularly use the PHQ-9 for MDS assessments revealed a set of common responses. When asked what they thought were the best ways to identify and address depression in residents, the majority responded that knowledge of the residents ($N = 8$) and direct observation of the residents ($N = 7$) remain the most important methods. In terms of experiences regarding the residents' ability to understand and respond to the PHQ-9 questions, many MDS clinician raters highlighted working with residents with dementia, confusion and cognitive impairment ($N = 8$) as well as the confusing nature or phrasing of certain questions ($N = 7$) as being obstacles of note. When asked how they address such obstacles, many replied that they try to accommodate the resident by taking more time to explain the questions or interviewing the residents at another

time ($N=8$) and when this is not at all possible, move on to conduct an informant interview with a staff member ($N=8$).

The majority of MDS clinician raters stated that they consult or discuss residents who trigger for depression with the rest of their team first and that they may at times make a psychiatric referral regardless of the PHQ-9 score ($N=8$). The most commonly stated reason for this was that there are many residents who seem depressed but do not trigger using the PHQ-9 questionnaire ($N=6$). When asked how staff members communicate with each other when it comes to assessing and treating depression in residents, almost all the MDS clinician raters replied that they consult regularly with other team members via e-mail, pager or conversation ($N=9$) in addition to weekly care plan meetings with the whole team including nurses, dieticians, therapists, social workers and physicians ($N=7$). When asked to identify barriers to such team coordination, many MDS clinician raters commented on communication issues like the loss of details between shifts ($N=5$), high caseloads that make it difficult to conduct quality assessments ($N=4$), and the need for better training of staff on dealing with dementia and depression ($N=3$).

4.3 | Quantitative results

Shown in Table 1 are the reliability estimates and ECVs from factor analyses of the PHQ-items for the Research and MDS assessments. The research PHQ-9 Cronbach's alpha internal consistency estimates ranged from 0.73 to 0.82 across ethnic/racial subgroups; the comparable coefficient range for the MDS assessment was somewhat lower, ranging from 0.63 to 0.76. The McDonald's omega reliability estimates were similar, 0.87 and 0.88 for the two assessments. There was evidence of slightly greater unidimensionality for the research assessment (ECV = 35.49) than for the MDS assessment (32.38).

Agreement: Shown in Table 2, Panel A, are the kappa coefficients examining agreement between the MDS clinician assessors and the standardised assessment of capacity in terms of determining capability for self-report assessment. There was moderate agreement between clinicians and researchers in determining resident capacity for self-report ($\kappa = 0.52$). MDS clinician assessors viewed a somewhat larger proportion of residents as capable of self-report than were so determined using the research capacity screen.

Shown in Table 2, Panel B, are the results for the trigger item, "Thoughts that you would be better off dead or of hurting yourself in some way." As shown, poor agreement in rating suicidal ideation was observed ($\kappa = 0.10$). Only two cases (1%) were identified by the MDS, whereas 15 (10%) were identified by the research PHQ-9. Although the overall agreement was relatively high, the sensitivity and positive predictive values were low.

As shown in Table 2, Panel C, there was no agreement beyond that expected by chance in classifying residents with at least moderate depression ($\kappa = -0.02$). The sensitivity and ppv were both very low. Given the marginal distributions, the observed kappa is much below the maximum obtainable kappa (0.67). The direction was consistent with the suicidal ideation item, with 13% of research PHQ-9 classifications above threshold as contrasted with 7% of MDS assessments.

The concordance between MDS and researcher ratings of depression in terms of total PHQ-9 severity scores, although significant, was low ($ICC = 0.22$, $p = .004$) (see Table 3). The mean score from the research assessments of 3.67 ($SD = 4.53$) was slightly higher than the mean score of 3.10 from the MDS ($SD = 3.86$; data not shown).

4.3.1 | Sensitivity analyses—The average number of days between assessments was 22.74 ($SD = 17.55$) days, with a median of 21 days; 75% of assessments were within 1 month. To account for the number of days between the MDS and researcher interviews as a potential control variable, the concordance analyses were conducted using different time points: (i) 0–14 days between assessments ($n = 53$), (ii) 0–30 days between assessments ($n = 111$) and (iii) full samples including 39 cases in which there were more than 30 days between the interviews. For the 0- to 14-day interval, the concordance between MDS and researcher ratings was low although significant for total PHQ-9 severity score ($ICC = 0.36$, $p = .003$). For suicidal ideation and classification of at least moderate depression, the agreements were fair ($\kappa = 0.37$ and 0.24 , respectively), and well below the maximum kappa (0.85) for classification of at least moderate depression (see Table 3).

For the 0- to 30-day interval, the concordance between clinician and researcher ratings was low although significant for the total PHQ-9 severity score ($ICC = 0.22$, $p = 0.007$). For suicidal ideation and diagnosis of at least moderate depression, the agreements were poor ($\kappa = 0.14$ and 0.02 , respectively).

In sensitivity analyses, residents for whom the informant interview was used by clinicians were removed (see Tables S1 and S2), and agreement examined for the 139 respondents assessed directly by both the researchers and MDS clinician raters. Slightly higher agreement indices were observed, which nevertheless remained poor ($ICC = 0.28$, $p < .001$). The mean score from the research protocol of 3.56 ($SD = 4.34$) was slightly higher than 2.88 ($SD = 3.71$) using the MDS (see Table S1). Additionally, the proportions scoring at the 10 or above threshold (cut-score) were similarly disparate: 10.8% using the research protocol and 5.8% using the MDS. The κ was 0.01 for classification of at least moderate depression (see Table S1). The agreement improved slightly but remained poor ($\kappa = 0.10$) for suicidal ideation (see Table S2).

4.3.2 | Subgroup analyses of Spanish speakers informed by interviews with clinicians—Many MDS clinician raters expressed that conducting interviews with Spanish-speaking residents was a problem. While Spanish-speaking staff ($N = 9$) were used as interpreters, there were challenges associated with their use, such as the translation of items into Spanish by the interpreter. As a result of these interviews, additional subgroup quantitative analyses examining results for residents interviewed in English and Spanish were performed. Shown in Tables S3 and S4 are the results of these analyses which show that eight of 19 cases (42%) identified by the research protocol as above threshold were Spanish speakers; none of whom were identified by the MDS clinician raters. Three of 10 persons (30%) identified as having depression by the MDS clinician raters were Spanish speaking; none of whom were so identified by the research protocol. Even more striking is that of the 15 individuals identified with suicidal ideation using the research protocol, seven were Spanish speakers (46%), none of whom were identified by the MDS. Of the two

identified by the MDS with suicidal ideation, both were English speakers. Overall, for the Spanish speakers, the research protocol identified 16% with PHQ-9 scores of 10 or greater, and 14% with thoughts of self-harm; this is in contrast to 6% and 0%, respectively, of the Spanish speakers assessed by the MDS clinician raters.

5 | DISCUSSION

Despite the documentation of high rates of depression in nursing homes (Abrams, Teresi, & Butin, 1992; Parmelee, Katz, & Lawton, 1992; Parmelee et al., 1989; Samuels & Katz, 1995; Teresi et al., 2001), the findings show under-recognition of depression using the MDS, particularly among Spanish speakers. The concordance between the total PHQ-9 scores assessed by the MDS and researchers was found to be low. While moderate agreement between the groups in determining resident capacity for self-report was observed, only poor to fair agreement for report of suicidal ideation and poor agreement for diagnosing at least moderate depression were observed. After omitting the cases for which MDS clinician raters used the staff informant interview, findings remained the same except that agreement for classification of moderate depression changed from none beyond chance to poor. A possible reason for such discordance may be that clinical staff members who were MDS clinician raters may have different approaches or attitudes towards using the PHQ-9 scale because of their different backgrounds and forms of training. This could in turn result in differences in the ways in which they rate a resident's depression even while using the same measure. This can occur even though the PHQ-9 is a standardised measure and assessors are instructed not to provide their own definitions of terms. Reported responses should be based on the residents' own interpretations, unshaped by whoever is administering the assessment (MED-PASS, Inc. 2009). In general, use of standardised interviews enhances depression recognition among staff (Teresi, Abrams, et al., 2000; Teresi et al., 2001).

Theoretically, there should be little difference between the MDS clinician raters and researchers in rating the degree of depression using the PHQ-9 in this sample of elders. However, because the researchers were trained professionals certified in administration of the measures and were provided with a script, specific coding instructions and a computer-assisted personal interview format, including professionally translated Spanish versions and Spanish-speaking assessors, a more standardised and controlled administration was expected. Consequently, improvements are needed in depression training and screening, and in particular, with the provision of consistent instructions as was provided in this study to the researchers.

There are still many challenges surrounding the use of the PHQ-9 in the nursing home setting. For example, the new MDS 3.0's increased volume of questions can be overwhelming or even frustrating to some nursing home residents and result in negative reactions to routine and often redundant interviews by staff. Additionally, definitions of specific phrases are not given. The items are intended to be administered as is without interpretation of terms (K. Kroenke, personal communication, January 5, 2010). Furthermore, administrators have not historically invested in psychosocial MDS training and training in the recognition of depression is an important activity to improve the recognition of depression in older people living in nursing homes (Smith, Stolder, Jagers, Liu, &

Haedtke, 2013). Embedding the PHQ-9 into the MDS 3.0 is a promising approach to promote better assessment and mental health care for nursing home residents. However, as discussed by Zimmerman, Connolly, Zlotnik, Bern-Klug, and Cohen (2012) regarding the MDS in general, that promise can be realised only if staff members have the necessary skills and clinical training to conduct more thorough assessments using the screens, then subsequently develop and implement appropriate care plans (Cohen et al., 2003; Zimmerman et al., 2012).

Suicide is one of the most serious consequences of depression, and this is particularly true among older adults who have among the highest completed suicide rate of any age group (Kramer et al., 2009). It is noted that admitting to suicidal ideation on the PHQ-9 may not be an indicator of actual suicide attempts (Schulberg et al., 2005). Suicide rates have historically been lower among nursing home residents than elders living in the community, but indirect self-harm and life-threatening behaviours such as refusing to eat or receive treatment are known to be common (Parmelee et al., 1989; Teresi et al., 2001). The PHQ-9 can facilitate the identification and referral of some residents who report suicidal ideation. Without the questionnaire, it is possible that many of these residents could be left unnoticed. However, the findings of this study showed poor consistency in the measurement of suicidal ideation, with potential underreporting in the MDS. A possible reason could be that residents refrained from expressing suicidal ideation to MDS clinician raters with whom they may have been familiar. For example, several MDS clinician raters commented during the in-depth interviews that some residents may underreport symptoms because they do not want to be any trouble for members of their family or nursing home staff.

A similar pattern of potential underreporting on the MDS was observed in terms of classifications of moderate depression, as indicated by a total severity score of 10 or higher. Scores of 10 or higher were more frequently observed using the research protocol. In instances where residents were directly interviewed while MDS clinician raters used staff informants were omitted, the degree of agreement improved slightly—from no agreement beyond that expected by chance to poor agreement—but the discordance remained nevertheless. Such disagreements could again potentially be explained by any or even all of the factors mentioned above including the interview setting, the fluctuation of mood with time, as well as the familiarity of the resident with the interviewer. As noted earlier, a score of 10 or above or a positive response on the suicide item triggers additional assessment. Although there is no evidence supporting such a contention for this study, in general there is the possibility of underreporting related to avoidance of additional assessments.

The findings were of relatively low agreement regardless of the time frame. Agreements did improve somewhat, although remained low and below the maximum kappa when the assessments were within 14 days of each other. It is possible that differences may be accounted for by the fact that the researchers and MDS clinician raters were administering the PHQ-9 as part of different protocols. The MDS clinician raters were interviewing the residents as part of their routine MDS 3.0 assessments; the researchers were interviewing the residents as part of an assessment battery that included several measures for depression as well as for functional status. Such differences in context may have induced certain residents

to respond differently to the researchers and MDS clinician raters even though they were being asked the same standardised set of questions.

During the in-depth interviews, the MDS clinician raters reported that the Spanish language barrier is a predominant obstacle when using the PHQ-9. Even when interpreters are available, many MDS clinician raters commented on how the use of interpreters itself can at times be challenging because they are not sure how the interpreter is wording the questions and worry that it may be in a way that affects how the resident responds. It is therefore important to note that 50 of the 155 resident interviews conducted by the researchers were in Spanish with Spanish-speaking assessors and a standardised translation of the PHQ-9. This model may have resulted in greater identification of depression. This finding has important implications for mental health disparities. The use of non-standardised protocols and non-Spanish-speaking assessors may result in severe underreporting of depression and suicidal ideation among Spanish speakers who will thus not be referred for further evaluation and treatment. Discordance in recognition of depression among minority as contrasted with other groups has been observed previously (Teresi et al., 2002). In that study, it was found that social workers as contrasted with nursing staff were more influenced in their ratings of depression by sociodemographic variables, such as African American status.

Item bias is a potential threat to the validity of self-reported data, and few analyses of differential item functioning (DIF) have been conducted comparing English- and Spanish-speaking respondents on the PHQ-9. Although one study supported the measurement equivalence of the PHQ-9 among English- and Spanish-speaking Hispanics (Merz, Malcarne, Roesch, Riley, & Sadler, 2011), another study found that items measuring anhedonia, sleep problems, appetite changes and guilt evidenced DIF for Hispanics (Huang, Chung, Kroenke, Delucchi, & Spitzer, 2006). It is possible that Hispanics over-report suicidal ideation for cultural reasons, unrelated to depression. However, expression of suicidal thoughts must be taken at face value as a potential severe indicator of depression and should be followed up for care planning. Referrals and treatments will not be offered if severe symptoms of depression are under-recognised.

5.1 | Limitations

A limitation of this study was the inability to control fully for the fluctuation of depressive symptoms over time. The majority (75%) were interviewed within a 1 month period, a look-back period used for many depression screens (Castro-Costa et al., 2008; Golden et al., 1984; Prince et al., 1999). The median number of days between assessments was 21 days, and the mean (after removal of outliers) was 20 days. To control partially for the time interval lag, the number of days between the MDS and research interviews was used as a control variable, and the concordance analyses were conducted using different time points. Doing so revealed that there was little agreement regardless of time frame although the agreement did improve somewhat when the assessments were within 14 days of each other, the look-back time frame for both the MDS and research PHQ-9 protocol.

There were other limitations to this project that should be addressed in future planned studies. These include the small sample size and restriction to a single institution and country, which may limit generalisability of findings. Despite these limitations, this is one of

the few studies examining convergence of ratings between locally collected PHQ-9 ratings and those using a research protocol. Moreover, this is one of the very few studies examining convergence by ethnic (language) group. Although there is inaccuracy in the measurement of race and ethnicity, for example LaVeist, 1994; examination of such subgroups remains important in reducing health and mental health disparities (National Research Council 2004).

6 | CONCLUSION

The findings of both the quantitative and in-depth interview components of this study support the conclusion that clinical assessments such as the PHQ-9, despite being previously validated for use in the nursing home, remain measures whose consistency and clinical utility require further investigation, particularly when administered to ethnically diverse groups who speak different languages.

7 | RELEVANCE TO CLINICAL PRACTICE

Identification of depression in older people living in nursing homes is essential to prevent feelings of isolation and loneliness (Haugan, Innstrand, & Moksnes, 2013). Knowledge of the residents and their direct observation were the most important methods identified by the MDS clinician raters to recognise depression. This implies that knowing residents well can improve identification of depression in residents. For example, the findings of Drageset, Eide, and Ranhoff (2011) support the practice of nursing home staff spending time with residents to be able to observe them for signs of depression. However, Mitchell and Kakkadasam (2011) concluded that depression recognition is less than optimal because nurses have difficulty identifying symptoms of depression in residents of nursing homes. Thus, use of standardised assessments was posited to increase depression recognition. In addition to training in administration of standardised depression assessments, it is recommended that training in depression recognition be provided to staff in long-term services and support settings. One such programme, developed for use with nursing staff (Abrams et al., 2016), includes three modules related to detection, treatment and assessment of depression, the latter of which provides specific training in administration of the PHQ-9.

This study has identified the difficulty that nursing home staff has in identifying depression, particularly in one ethnically diverse group of nursing home residents. This means that many older people living in nursing homes could have unidentified and untreated depression in the final years of their life. While nursing home staff believe that by knowing a resident they can identify when that resident becomes depressed, without the use of standardised depression assessment tools, many incidents of depression will be missed.

It is recommended highly that adequate assessment of depression in nursing home residents from ethnically diverse groups become a high priority for all nursing homes. Accurate assessment requires identification of appropriate ethnically diverse staff trained comprehensively in using clinical assessments such as the PHQ-9. Furthermore, standardised translations of these instruments should be made available to staff. The use of a

standardised assessment tool would help staff gain knowledge and expertise in identifying depression.

Overall, the PHQ-9 has the potential to identify residents in need of further evaluation by providing quantitative evidence to care staff that depressive symptomatology is such that appropriate interventions are recommended (Taylor, 2014). Use of an instrument that performs well among older people, including those from different ethnic and language groups, is important for accurate depression recognition and subsequent intervention (Drageset et al., 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding information

Funding for this study conducted by the Consortium of New York Geriatric Education Centers was from the Health Resources and Services Administration (J. Howe, PI), Grant number 5UB4 HP19194-05. Partial funding for the secondary data analyses was provided by the National Institute on Aging (NIA)-funded Mt. Sinai Pepper Center (A. Siu, Mount Sinai, P30-AG028741) and the NIA-funded Cornell Roybal Center, Translational Research Institute on Pain in Later Life-II (M.C. Reid, PI, Cornell, 5P30AG022845)

References

- Abrams R, Nathanson M, Silver S, Ramirez M, Toner J, Teresi JA. A training program to enhance recognition of depression in nursing homes, assisted living and other long-term care settings: Description and evaluation. *Gerontology & Geriatrics Education*. 2016; 38:325–345. [PubMed: 26885893]
- Abrams RC, Teresi JA, Butin D. Depression in nursing home residents. *Clinics in Geriatric Medicine*. 1992; 8(2):309–322. [PubMed: 1600481]
- Bell M, Goss AJ. Recognition, assessment and treatment of depression in geriatric nursing home residents. *Clinical Excellence for Nurse Practitioners*. 2001; 5(1):26–36. [PubMed: 11154391]
- Brown MN, Lapane KL, Luisi AF. The management of depression in older nursing home residents. *Journal of the American Geriatric Society*. 2002; 50(1):69–76.
- Castro-Costa E, Dewey M, Stewart R, Banerjee S, Huppert F, Mendonca-Lima C, ... Prince M. Ascertaining late-life depressive symptoms in Europe: An evaluation of the survey version of the EURO-D scale in 10 nations. The SHARE project. *International Journal of Methods in Psychiatric Research*. 2008; 17:12–29. [PubMed: 18286461]
- Centers for Disease Control and Prevention (CDC). Surveillance for violent deaths – National violent death reporting system, 16 states, 2007. *MMWR Morbidity and Mortality Weekly Report*. 2010; 59:1–50. [PubMed: 20075837]
- Chen LY, Liu LK, Hwang AC, Lin MH, Peng LN, Chen LK, ... Chang PL. Impact of malnutrition on physical, cognitive function and mortality among older men living in veteran homes by minimum data set: A prospective cohort study in Taiwan. *Journal of Nutrition, Health & Aging*. 2016; 20(1): 41–47.
- Chung H, Teresi J, Guarnaccia P, Meyers B, Holmes D, Bobrowitz T, ... Ferran E Jr. Depressive symptoms and psychiatric distress in low income Asian and Hispanic primary care patients: Prevalence and recognition. *Community Mental Health Journal*. 2003; 39(1):33–46. [PubMed: 12650554]
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.

- Cohen CI, Hyland K, Kimhy D. The utility of mandatory depression screening of dementia patients in nursing homes. *American Journal of Psychiatry*. 2003; 160(11):2012–2017. [PubMed: 14594749]
- Connolly RP, Downes D, Reuter J, Fogler S. M.D.S. 3.0: More training needed: Social work concerns about interview requirements. *Provider*. 2010; 36(10):85–89. [PubMed: 21038823]
- Creswell, JW. *Research design: Qualitative, quantitative, and mixed methods approaches*. 2. Thousand Oaks, CA: Sage Publications Inc; 2003.
- Delaney C, Fortinsky R, Doonan L, Grimes RLW, Terra-Lee P, Rosenberg S, Bruce M. Depression screening and interventions for older home health care patients: Program design and training outcomes for a train-the-trainer model. *Home Health Care Management and Practice*. 2011; 23:435–445.
- Drageset J, Eide GE, Ranhoff AH. Depression is associated with poor functioning in activities of daily living among nursing home residents without cognitive impairment. *Journal of Clinical Nursing*. 2011; 20:3111–3118. <https://doi.org/10.1111/j.1365-2702.2010.03663.x>. [PubMed: 21592245]
- Golden RR, Teresi JA, Gurland BJ. Detection of dementia and depression cases with the Comprehensive Assessment and Referral Evaluation Interview Schedule. *International Journal of Aging and Human Development*. 1983; 16:242–254.
- Golden RR, Teresi JA, Gurland B. Development of indicator scales for the comprehensive assessment and referral evaluation (CARE) interview schedule. *Journal of Gerontology*. 1984; 39(2):138–146. [PubMed: 6699367]
- Gurland B, Wilder D. The CARE interview revisited: Development of an efficient, systematic, clinical assessment. *Journal of Gerontology*. 1984; 39:129–137. [PubMed: 6699366]
- Gurland, BJ., Wilder, D., Lantigua, R., Mayeux, R., Stern, Y., Chen, J., ... Killeffer, E. Differences in rates of dementia between ethno-racial groups. In: Martin, LG., Soldo, BJ., editors. *Racial and ethnic differences in the health of older Americans*. Washington, DC: National Academy Press; 1997. p. 233–269.
- Haugan G, Innstrand ST, Moksnes UK. The effect of nurse-patient interaction on anxiety and depression in cognitively intact nursing home patients. *Journal of Clinical Nursing*. 2013; 22:2192–2205. <https://doi.org/10.1111/jocn.12072>. [PubMed: 23472821]
- Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal of General Internal Medicine*. 2006; 21(6):547–552. [PubMed: 16808734]
- Jongenelis K, Pot AM, Eisses AM, Beekman AT, Kluiters H, Ribbe MW. Prevalence and risk indicators of depression in elderly nursing home patients: The AGED study. *Journal of Affective Disorders*. 2009; 83(2–3):135–142.
- Kerber CS, Dyck MJ, Culp KR, Buckwalter K. Comparing the Geriatric Depression Scale, minimum data set, and primary care provider diagnosis for depression in rural nursing home residents. *Journal of the American Psychiatric Nurses Association*. 2005; 11:269–275.
- Kramer D, Allgaier AK, Fejtikova S, Mergl R, Hegerl U. Depression in nursing homes: Prevalence, recognition, and treatment. *International Journal of Psychiatry in Medicine*. 2009; 39(4):345–358. [PubMed: 20391857]
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001; 16(9):606–613. [PubMed: 11556941]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- LaVeist TA. Beyond dummy variables and sample selection: What health services researchers ought to know about race as a variable. *Health Services Research*. 1994; 29(1):1–16. [PubMed: 8163376]
- Mann A, Graham N, Ashby D. Psychiatric illness in residential homes for the elderly: A survey in one London borough. *Age and Ageing*. 1984; 13(5):257–265. [PubMed: 6496236]
- McCabe MP, Mellor D, Davison TE, Hallford DJ, Goldhammer DL. Detecting and managing depressed patients: Palliative care nurses' self-efficacy and perceived barriers to care. *Journal of Palliative Medicine*. 2012; 15(4):463–467. [PubMed: 22475193]
- MED-PASS, Inc. (Version 3.0). *Comprehensive user manual resident assessment instrument section D: Mood*. Miamisburg, OH: MED-PASS, Inc; 2009.

- Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity and Ethnic Minority Psychology*. 2011; 17(3):309–316.
- Mitchell AJ, Kakkadasam V. Ability of nurses to identify depression in primary care secondary care and nursing homes – A meta-analysis of routine clinical accuracy. *International Journal of Nursing Studies*. 2011; 48:359–368. <https://doi.org/10.1016/j.ijnurstu.2010.05.012>. [PubMed: 20580001]
- Morris JN, Hawes C, Fries BE, Phillips C, Mor V, Katz S, ... Friedlob AS. Designing the national resident instrument for nursing homes. *The Gerontologist*. 1990; 30:293–307. [PubMed: 2354790]
- National Research Council. Measuring racial discrimination. Washington, DC: The National Academies Press; 2004. Panel on methods for assessing discrimination. Committee on National Statistics Division of Behavioral and Social Sciences and Education
- Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: Assessment and prevalence estimation. *Journals of Gerontology: Medical Sciences*. 1989; 44(1):M22–M29.
- Parmelee PA, Katz IR, Lawton MP. Incidence of depression in long-term care settings. *Journals of Gerontology: Medical Sciences*. 1992; 47(6):M189–M196.
- Pratt LA, Brody DJ. Depression in the U.S. household population 2009–2012. NCHS Data Brief. 2014; 172:1–8.
- Prince MJ, Reischies F, Beekman AT, Fuhrer R, Jonker C, Kivela SL, ... Copeland JR. Development of the EURO-D scale –A European, Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*. 1999; 174:330–338. [PubMed: 10533552]
- Sales AE, Boström AM, Bucknall T, Draper KZ, Fraser K, Schalm C, Warren S. The use of data for process and quality improvement in long term care and home care: A systematic review of the literature. *The Journal of the American Medical Directors Association*. 2012; 13(2):103–113. [PubMed: 21450243]
- Saliba D., Buchanan, J. Final Report to the Quality Measurement and Health Assessment Group Office of Clinical Standards and Quality, Centers for Medicare and Medicaid Services. 2008. Development and validation of a revised nursing home assessment: MDS 3.0.
- 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. *The Journal of the American Medical Directors Association*. 2012; 13(7):618–625. [PubMed: 22796361]
- Samuels S, Katz I. Depression in the nursing home. *Psychiatric Annals*. 1995; 25(7):419–424.
- Schulberg HC, Lee PW, Bruce ML, Raue PJ, Lefever JJ, Williams JW Jr, ... Nutting PA. Suicidal ideation and risk levels among primary care patients with uncomplicated depression. *Annals of Family Medicine*. 2005; 3(6):523–528. [PubMed: 16338916]
- Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. *Psychological Medicine*. 2006; 36(1):27–36. [PubMed: 16202189]
- Simons K, Connolly RP, Bonifas R, Allen PD, Bailey K, Downes D, Galambos C. Psychosocial assessment of nursing home residents via MDS 3.0: Recommendations for social service training, staffing, and roles in interdisciplinary care. *The Journal of the American Medical Directors Association*. 2012; 13:190e9–190.e15.
- Smith M, Stolder ME, Jagers B, Liu MF, Haedtke C. Depression training in nursing homes: Lessons learned from a pilot study. *Issues in Mental Health Nursing*. 2013; 34:90–102. <https://doi.org/10.3109/01612840.2012.723798>. [PubMed: 23369120]
- Tarride JE, Burke N, Leslie WD, Morin SN, Adachi JD, Papaioannou A, ... Hopkins RB. Loss of health related quality of life following low-trauma fractures in the elderly. *BMC Geriatrics*. 2016; 16(84):1–11. <https://doi.org/10.1186/s12877-016-0259-5>. [PubMed: 26729190]
- Taylor WD. Depression in the elderly. *New England Journal of Medicine*. 2014; 371(13):1228–1236. [PubMed: 25251617]
- Teresi, JA., Abrams, R., Holmes, D. Measurement of depression and depression recognition among individuals with cognitive impairment. In: Albert, S., Logsdon, R., editors. *Assessing quality of life in Alzheimer's disease*. New York, NY: Springer Publishing Company; 2000. p. 127-151.
- Teresi JA, Abrams R, Holmes D, Ramirez M, Eimicke J. Prevalence and depression recognition in nursing homes. *Social Psychiatry Psychiatric Epidemiology*. 2001; 36(12):613–620. [PubMed: 11838834]

- Teresi JA, Abrams R, Holmes D, Ramirez M, Shapiro C, Eimicke J. The influence of cognitive impairment, illness, gender and African-American status on psychiatric ratings and staff recognition of depression. *American Journal of Geriatric Psychiatry*. 2002; 10(5):506–514. [PubMed: 12213684]
- Teresi JA, Golden R, Gurland B. Concurrent and predictive validity of the indicator-scales developed for the Comprehensive Assessment and Referral Evaluation (CARE) interview. *Journal of Gerontology*. 1984; 39(2):158–165. [PubMed: 6699369]
- Teresi JA, Kleinman M, Ocepek-Welikson K, Ramirez M, Gurland B, Lantigua R, Holmes D. Applications of item response theory to the examination of the psychometric properties and differential item functioning of the CARE Dementia Diagnostic Scale among samples of Latino, African-American and White non-Latino Elderly. *Research on Aging*. 2000; 22:738–773.
- Wilder D, Cross P, Chen J, Gurland B, Lantigua R, Teresi J, ... Encarnacion P. Operating characteristics of brief screens for dementia in a multicultural population. *The American Journal of Geriatric Psychiatry*. 1995; 3:1–12. [PubMed: 28530953]
- Zimmerman S, Connolly R, Zlotnik JL, Bern-Klug M, Cohen LW. Psychosocial care in nursing homes in the era of the MDS 3.0: Perspectives of the experts. *Journal of Gerontological Social Work*. 2012; 55(5):444–461. [PubMed: 22783960]

What does this paper contribute to the wider global clinical community?

- Study results suggest that clinical assessments such as the PHQ-9, despite being previously validated for use in long-term care facilities, remain measures whose consistency and clinical utility require further investigation, particularly when administered to ethnically diverse groups who speak languages other than English.
- Although standardised assessment measures have been developed, insufficient staff training remains a problem and there persists the need for research on how practitioners can more effectively use these assessments and the resulting data for referral and treatment.
- This study demonstrates that depression recognition in long-term care facilities, particularly among residents from ethnically diverse groups who speak languages other than English, remains a challenge that has implications for the implementation of quality of care improvement strategies.

TABLE 1

Reliability estimates for the PHQ-9: Cronbach's alpha internal consistency across ethnic/racial subgroups and McDonald's omega total for the research and MDS assessment

	Cronbach's alpha	Explained common variance (eigenvalues) from factor analyses	McDonald's omega total
RA PHQ-9			
Total sample ($n = 150$)	0.75	35.49 (3.19, 1.10)	0.87
Black, non-Hispanic ($n = 62$)	0.73		
Hispanic ($n = 57$)	0.73		
White, non-Hispanic ($n = 27$)	0.82		
MDS PHQ-9			
Total sample ($n = 153$)	0.72	32.38 (2.91, 1.34, 1.21)	0.88
Black, non-Hispanic ($n = 62$)	0.76		
Hispanic ($n = 58$)	0.71		
White, non-Hispanic ($n = 28$)	0.63		

PHQ-9, 9-item Patient Health Questionnaire; RA, researcher; MDS, Minimum Data Set.

TABLE 2

Comparisons of MDS clinician raters versus research findings

Panel A: resident capacity for self-report (N = 227)					κ = 0.52, p = .062; moderate agreement between the groups	
Researcher						
No self-report		Self-report		Total		
MDS clinician ratings	No self-report	Count	43	16	59	
		Row %	72.88	27.12	100.00	
	Self-report	Count	29	139	168	
		Row %	17.26	82.74	100.00	
Total		Count	72	155	227	
		Row %	31.72	68.28	100.00	
Panel B: residents with suicidal ideation (N = 148) ^{a,b}					κ = 0.10, p = .060; poor agreement between the groups	
Researcher						
No suicidal ideation		Suicidal ideation		Total		
MDS clinician raters	No suicidal ideation	Count	132	14	146	
		Row %	90.41	9.59	100.00	
		Column %	99.25	93.33	98.65	
	Suicidal ideation	Count	1	1	2	
		Row %	50.00	50.00	100.00	
		Column %	0.75	6.67	1.35	
Total		Count	133	15	148	
		Row %	89.86	10.14	100.00	
		Column %	100.00	100.00	100.00	
				Maximum Kappa	0.22	
					Sensitivity	0.07
					Specificity	0.99
					PPV	0.50
					NPV	0.90
					OCC	0.90
					PEA	0.89
					OR (95% CI)	9.43 (0.56, 159.14)
					Kappa	0.10
					Maximum Kappa	0.22
Panel C: residents with at least moderate depression (score of 10+) (N = 150) ^a					κ = −0.02, p = .793; no agreement beyond that predicted by chance	
Researcher						
Score of less than 10		Score of 10 or greater		Total		
MDS clinician raters	Score of less than 10	Count	122	18	140	
		Row %	87.14	12.86	100.00	
		Column %	93.13	94.74	93.33	
	Score of 10 or greater	Count	9	1	10	
					NPV	0.87
					Sensitivity	0.05
					Specificity	0.93
					PPV	0.10

Researcher		κ = 0.52, p = .062; moderate agreement between the groups	
		No self-report	Self-report
Panel A: resident capacity for self-report (N = 227)		Total	Total
Row %	90.00	100.00	OCC
Column %	6.87	6.67	PEA
Count	131	150	OR (95% CI)
Row %	87.33	100.00	Kappa
Column %	100.00	100.00	Maximum Kappa
			0.67

MDS, Minimum Data Set; PPV, positive predictive value; NPV, negative predictive value; OCC, overall correct classification (percent agreement); PEA, percent expected agreement; OR, odds ratio.

^aResearcher is treated as the gold standard.

^bItem is “Thoughts that you would be better off dead or of hurting yourself in some way?”.

Concordance analyses between researchers and MDS clinician raters using different time points (intraclass correlation coefficients and kappa values)^a

TABLE 3

		0–14 days between assessments (<i>n</i> = 53)			0–30 days between assessments (<i>n</i> = 111)			Full samples (<i>n</i> = 150)		
		ICC/Kappa	<i>p</i> -Value	Maximum Kappa	ICC/Kappa	<i>p</i> -Value	Maximum Kappa	ICC/Kappa	<i>p</i> -Value	Maximum Kappa
PHQ-9 total score		0.36	.003	–	0.22	.007	–	0.22	.004	–
PHQ-9 score	10	0.24	.082	0.85	0.02	.827	0.67	30.02	.793	0.67
Suicidal ideation		0.37	.006	0.48	0.14	.042	0.31	0.10	.060	0.22

ICC, intraclass correlation coefficients; MDS, Minimum Data Set; PHQ-9, 9-item Patient Health Questionnaire.

The mean is 22.74 (*SD* = 17.55) days between assessments; median = 21 days; min = 0; max = 83. Excluding eight outliers, the mean is 20.06 (*SD* = 13.53); max = 57 days.

^aNo statistics were computed for the 15–30 days between assessments subgroup because the MDS reported Suicidal Ideation variable was constant.