NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0690 **NQF Project:** Nursing Homes 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Percent of Residents Who Have Depressive Symptoms (Long-Stay)
De.2 **Brief description of measure***:* This measure is based on data from MDS 3.0 assessments of nursing home residents. Either a resident interview measure or a staff assessment measure will be reported. The preferred version is the resident interview measure. The resident interview measure will be used unless either there are three or more missing sub-items needed for calculation or the resident is rarely or never understood, in which cases the staff assessment measure will be calculated and used. These measures use those questions in MDS 3.0 that comprise the Patient Health Questionnaire (PHQ-9) depression instrument. The PHQ-9 is based on the diagnostic criteria for a major depressive disorder in the DSM-IV.
1.1-2 **Type of Measure:** Outcome De.3 **If included in a composite or paired with another measure, please identify composite or paired measure**
De.4 **National Priority Partners Priority Area:** Population health De.5 **IOM Quality Domain**: Patient-centered De.6 **Consumer Care Need:**

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary	**A** **Y**☐ **N**☐

A.4 Measure Steward Agreement attached:	
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public Reporting	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** No, testing will be completed within 24 months D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** **Staff Reviewer Name(s):**	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Depression is a very expensive,complicating, and treatable factor for nursing facility residents. The total economic cost of depression in the U.S. in CY 2000 was $83.1 billion, including $26.1 billion in direct medical costs.(1) In the nursing facility environment, depression can be triggered by a number of elements of physical or cognitive decline, and by the circumstances of nursing home residence itself (in addition to other causes), but can be under diagnosed and under treated.(2) As summarized by Saliba and Buchanan: Research conducted before the national implementation of the MDS demonstrated that the prevalence of major depression among cognitively intact or moderately impaired nursing facility residents was 20-25%. In addition, another 30% of residents had less severe, but nevertheless clinically significant depression.(3) However . . . only about 10% of residents with recognized depression were treated.(4) More recent studies reveal that, despite an emphasis on depression in the MDS and associated quality indicators, as well as an almost 3 fold increase in the number of residents prescribed antidepressants,(5) 34% of residents may have clinically significant depressive symptoms.(6)	**1a** C☐ P☐ M☐ N☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [KP1]: 1a. The measure focus addresses:
- a specific national health goal/priority identified by NQF's National Priorities Partners; OR
- a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

For the second quarter of 2008, the current measure ("Percent of Residents Who Have Become More Depressed or Anxious") based on MDS 2.0 data averages 14.9% nationally, with statewide averages ranging from 9.2% to 30%. Therefore, depression among the nursing home residents is a significant clinical issue.

1a.4 **Citations for Evidence of High Impact:** 1. Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, Corey-Lisle PK. The economic burden of depression in the United States: how did it change between 1990 and 2000. J Clin Psychiatr. 2003;64(12):1465-75.

2. Simmons SF, Cadogan MP, Cabrera GR, et al. The Minimum Data Set depression quality indicator: does it reflect differences in care processes? Gerontologist. 2004;44:554-64.

3. Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: assessment and prevalence estimation. J Gerontol. 1989;44(1):M22-9.

4. Heston LL, Garrard J, Makris L, et al. Inadequate treatment of depressed nursing home elderly. J Am Geriatr Soc. 1992;40(11):1117-22.

5. Weintraub D, Datto CJ, Streim JE, et al. Second-generation issues in the management of depression in nursing homes. J Am Geriatr Soc. 2002;50(12):2100-1; author reply, 2101.

6. Datto CJ, Oslin DW, Streim JE, et al. Pharmacologic treatment of depression in nursing home residents: a mental health services perspective. J Geriatr Psychiatry Neurol. 2002;15(3):141-6.

1b. Opportunity for Improvement

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Facilities can use information from this measure to identify the extent to which their long-stay residents have symptoms of major depression and to develop quality improvement initiatives to ensure adequate care planning. Fewer residents with symptoms of major depression and fewer symptoms of major depression in individual residents are the expected quality improvements.

1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**
Depression in short-stay residents may result from a number of temporary factors based on recent acute inpatient admission that may recede with time, and treatment for depression may not have had sufficient time to become effective.

Depression is a treatable condition. As summarized by Saliba and Buchanan:

Research conducted before the national implementation of the MDS demonstrated that the prevalence of major depression among cognitively intact or moderately impaired nursing facility NH residents was 20-25%. In addition, another 30% of residents had less severe, but nevertheless clinically significant depression.(1) However . . . only about 10% of residents with recognized depression were treated.(2) More recent studies reveal that, despite an emphasis on depression in the MDS and associated quality indicators, as well as an almost 3 fold increase in the number of residents prescribed antidepressants,(3) 34% of residents may have clinically significant depressive symptoms.(4)

The MDS 3.0 items are expected to better identify depression in the nursing facility populations than the MDS 2.0 items. Much of the following information presented in this section is taken from Saliba and Buchanan (2008), which presents the results of the CMS-initiated project to create version 3.0 of the MDS.(5) In that project, the research team engaged in an iterative process to incorporate provider and consumer input, expert consultation, scientific advances in clinical knowledge about screening and assessment, CMS experience, and intensive item development and testing. This process resulted in national testing of MDS 3.0 in 71 community nursing facilities and 19 Veterans Affairs (VA) nursing homes. The national test directly examined: agreement between assessors (reliability); validity of new items; response rates for interview items; user satisfaction and feedback on changes; time to complete the assessment; and comparison of item distributions between MDS 3.0 and MDS 2.0.

This MDS 3.0 Development Project oversampled short-stay residents (those who are discharged within 100 days

1b
C ☐
P ☐
M ☐
N ☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

of admission). It is therefore likely that their sample includes fewer residents with serious cognitive impairment than would typically be present in the long-stay nursing facility population.(5)

Saliba and Buchanan note the following about MDS 2.0:

- The current MDS 2.0 list of 15 observed indicators of depression has poor sensitivity for identifying persons with depressive symptoms or depression.(6, 7, 8, 9, 10, 11)
- A consensus statement from the American Geriatrics Society (AGS) and the American Association for Geriatric Psychiatry (AAGP) concluded that the MDS alone, as currently used, is not adequate for depression screening and recommended that additional instruments be used.(12)
- Only 22% of nurses in their survey reported that the MDS 2.0 mood items are easy to complete accurately.(5)

Concerns about the currently used measure focus on two areas. The first is that screening specifically for depression would be valuable and that anxiety and depression should not be collapsed into a single construct. The second is that some of these indicators may have causes unrelated to depression or anxiety. In particular, negative statements, repetitive verbalizations, crying, tearfulness, and repetitive physical movements may result from other factors, such as bereavement or cognitive impairment. Also, leaving food uneaten may be caused in part by Federal regulations related to meal frequency and nutritional adequacy, which lead many nursing homes to be reluctant to allow residents to select their own food portions; as a result, residents may leave food uneaten because the portions provided are too large.

1b.3 Citations for data on performance gap:

1. Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: assessment and prevalence estimation. J Gerontol. 1989;44(1):M22-9.

2. Heston LL, Garrard J, Makris L, et al. Inadequate treatment of depressed nursing home elderly. J Am Geriatr Soc. 1992;40(11):1117-22.

3. Weintraub D, Datto CJ, Streim JE, et al. Second-generation issues in the management of depression in nursing homes. J Am Geriatr Soc. 2002;50(12):2100-1; author reply, 2101.

4. Datto CJ, Oslin DW, Streim JE, et al. Pharmacologic treatment of depression in nursing home residents: a mental health services perspective. J Geriatr Psychiatry Neurol. 2002;15(3):141-

5. Saliba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

6. Teresi J, Abrams R, Holmes D, et al. Prevalence of depression and depression recognition in nursing homes. Soc Psychiatry Psychiatr Epidemiol. 2001;36(12):613-20.

7. Anderson RL, Buckwalter KC, Buchanan RJ, et al. Validity and reliability of the Minimum Data Set Depression Rating Scale (MDSDRS) for older adults in nursing homes. Age Ageing. 2003;32(4):435-8.

8. Horgas AL, Tsai PF. Analgesic Drug prescription and use in cognitively impaired nursing home residents. Nurs Res. 1998;47(4):235-42.

9. McCurren C. Assessment for depression among nursing home elders: evaluation of the MDS mood assessment. Geriatric Nurs. 2002;23(2):103-8.

10. Lawton MP, Casten R, Parmelee PA, et al. Psychometric characteristics of the Minimum Data Set II: validity. J Am Geriatr Soc. 1998;46(6):736-44.

11. 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.

12. American Geriatrics Society and American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral symptoms associated with dementia. J Am Geriatr Soc. 2003;51(9):1287-98.

1b.4 Summary of Data on disparities by population group:
Several studies have analyzed racial differences in depression in nursing homes. One study found that African American residents were less likely to be diagnosed and less likely to receive treatment.(1) Another study also found African Americans were less likely to be diagnosed, but found no significant racial differences in recorded mood or behavior symptomatology or in the pharmacologic treatment of mental illness.(2) Among community-dwelling older persons, studies have largely shown a greater incidence of depression and depressive symptoms in blacks than in whites, although some of this difference is due to intervening socioeconomic factors.(3, 4, 5 ,6) However, some studies found lower rates of depression for blacks than for whites and Hispanics, or no differences, including one study of nursing home residents.(7, 8, 9)

1b.5 Citations for data on Disparities:

1. Levin CA, Wei W, Akincigil A, Lucas JA, Bilder S, Crystal S. Prevalence and treatment of diagnosed depression among elderly nursing home residents in Ohio. J Am Med Dir Assoc. 2007 Nov;8(9):585-94. Epub 2007 Oct 22.

2. Zisselman M, Smith R, Smith S, Daskalakis C, Sanchez F. Racial and socioeconomic differences in psychiatric symptoms in nursing home residents: a Minimum Data Set–based pilot study. J Am Med Dir Assoc. 2006;7(1):17-22.

3. Cohen CI, Magai C, Yaffee R, Walcott-Brown L. Racial differences in syndromal and subsyndromal depression in an older urban population. Psychiatr Serv. 2005 Dec;56(12):1556-63.

4. Dunlop DD, Song J, Lyons JS, Manheim LM, Chang RW. Racial/ethnic differences in rates of depression among preretirement adults. Am J Public Health. 2003 Nov;93(11):1945-52.

5. Sachs-Ericsson N, Plant EA, Blazer DGRacial differences in the frequency of depressive symptoms among community dwelling elders: the role of socioeconomic factors. Aging Ment Health. 2005 May;9(3):201-9.

6. Skarupski KA, Mendes de Leon CF, Bienias JL, Barnes LL, Everson-Rose SA, Wilson RS, Evans DA. Black-white differences in depressive symptoms among older adults over time. J Gerontol B Psychol Sci Soc Sci. 2005 May;60(3):P136-42.

7. Cohen CI, Hyland K, Magai C. Depression among African American nursing home patients with dementia. Am J Geriatr Psychiatry. 1998 Spring;6(2):162-75.

8. Fyffe DC, Sirey JA, Heo M, Bruce ML. Late-life depression among black and white elderly homecare patients. Am J Geriatr Psychiatry. 2004 Sep-Oct;12(5):531-5.

9. Steffens DC, Fisher GG, Langa KM, Potter GG, Plassman BL. Prevalence of depression among older Americans: the Aging, Demographics and Memory Study. Int Psychogeriatr. 2009 Oct;21(5):879-88. Epub 2009 Jun 12.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** This measure of depressed residents is directly related to an important component of overall health status. The illness benefits from treatment and untreated depression may contribute to a resident's decline. Better screening will increase the likelihood of treatment. Saliba and Buchanan report that 84 percent of the nurses in their study felt that the interview could inform facility care plans, and that 86 percent reported that even in the limited number of residents assessed, the interview items provided new insights into resident mood.(1) They also reported that for the PHQ-9 staff observation version, 90% felt that staff detection and communication about mood disorder might improve if they learned to watch for these signs and symptoms. This is an important finding, as Lapid and Rummans (2003) reported that geriatric depression is a common but frequently unrecognized or inadequately treated condition in the elderly population.

1c C ☐ P ☐ M ☐ N ☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Treatment can be effective for depression, although there are particular challenges among seniors. According to Barkin et al. (2000), antidepressant pharmacotherapy combined with cognitive and behavioral therapy appears to offer the most benefit to the patient. However, the elderly present pharmacotherapeutic challenges in terms of the absorption, distribution, metabolism, and elimination of most antidepressants. Taking other medicines may act to induce or inhibit the metabolism of the antidepressant that the patient is using.

Reynolds et al. (2006) note that elderly patients with major depression, including those having a first episode, are at high risk for recurrence of depression, disability, and death, and tested the efficacy of maintenance treatment with an selective serotonin reuptake inhibitor (SSRI) and monthly interpersonal psychotherapy in patients 70 years of age or older who had depression in a randomized, double-blind, placebo-controlled trial. They concluded that patients with major depression who had a response to initial treatment with paroxetine and psychotherapy were less likely to have recurrent depression if they received two years of maintenance therapy with the SSRI. However, monthly maintenance psychotherapy did not prevent recurrent depression.

Pinquart et al. (2006) performed a meta-analysis comparing pharmacotherapy and psychotherapy treatments for later-life depressive conditions. They found that available treatments for depression work, with effect sizes that are moderate to large, and concluded that because psychotherapy and pharmacotherapy did not show strong differences in effect sizes, treatment choice should be based on other criteria, such as contraindications, treatment access, or patient preferences.

1. Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, Corey-Lisle PK. The economic burden of depression in the United States: how did it change between 1990 and 2000. J Clin Psychiatr. 2003;64(12):1465-75.
2. Simmons SF, Cadogan MP, Cabrera GR, et al. The Minimum Data Set depression quality indicator: does it reflect differences in care processes? Gerontologist. 2004;44:554-64.
3. Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: assessment and prevalence estimation. J Gerontol. 1989;44(1):M22-9.
4. Heston LL, Garrard J, Makris L, et al. Inadequate treatment of depressed nursing home elderly. J Am Geriatr Soc. 1992;40(11):1117-22.
5. Weintraub D, Datto CJ, Streim JE, et al. Second-generation issues in the management of depression in nursing homes. J Am Geriatr Soc. 2002;50(12):2100-1; author reply, 2101.
6. Datto CJ, Oslin DW, Streim JE, et al. Pharmacologic treatment of depression in nursing home residents: a mental health services perspective. J Geriatr Psychiatry Neurol. 2002;15(3):141-6.

1c.2-3. Type of Evidence: Randomized controlled trial, Observational study

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
The MDS 3.0 contains a resident interview version and a staff assessment version of the PHQ-9, which is based on the diagnostic criteria for a major depressive disorder in the DSM-IV. This targeted focus on major depression is a significant shift from the MDS 2.0 data currently used in the quality indicator measure, "Percent of Long-Stay Residents Who Are More Depressed or Anxious." The current measure (being replaced by the proposed measure) from the MDS 2.0 data combines two separate conditions (depression and anxiety), as well as situations that may result from other causes entirely: distress, crying/tearfulness, motor agitation, leaves food uneaten, repetitive health complaints, repetitive/recurrent verbalizations, negative statements, and mood symptoms not easily altered.

With the focus on major depression, Saliba and Buchanan report that 84% of the nurses in their study believed that the interview could inform facility care plans, and that 86% reported that even in the limited number of residents assessed, the interview items provided new insights into resident mood.(1) They also reported that for the PHQ-9 staff observation version, 90% of the nurses believed that staff detection and communication about mood disorder might improve if they learned to watch for these signs and symptoms. This is an important finding, as Lapid and Rummans (2003) reported that geriatric depression is a common but frequently unrecognized or inadequately treated condition in the elderly population. Both major and minor depression are associated with high mortality rates if left untreated.(2)

There are particular challenges in treating depression among seniors. According to Barkin et al. (2000), antidepressant pharmacotherapy combined with cognitive and behavioral therapy appears to offer the most benefit to the patient.(3) However, the elderly present pharmacotherapeutic challenges in terms of the absorption, distribution, metabolism, and elimination of most antidepressants. Taking other medicines may act to induce or inhibit the metabolism of the antidepressant that the patient is using.

Reynolds et al. (2006) note that elderly patients with major depression, including those having a first episode, are at high risk for recurrence of depression, disability, and death, and tested the efficacy of maintenance treatment with an SSRI and monthly interpersonal psychotherapy in patients 70 years of age or older who had depression in a randomized, double-blind, placebo-controlled trial.(4) They concluded that patients with major depression who had a response to initial treatment with paroxetine and psychotherapy were less likely to have recurrent depression if they received two years of maintenance therapy with the SSRI. However, monthly maintenance psychotherapy did not prevent recurrent depression.

Pinquart et al. (2006) performed a meta-analysis comparing pharmacotherapy and psychotherapy treatments for later-life depressive conditions.(5) They found that available treatments for depression work, with effect sizes that are moderate to large, and concluded that because psychotherapy and pharmacotherapy did not show strong differences in effect sizes, treatment choice should be based on other criteria, such as contraindications, treatment access, or patient preferences.

1. Saliba D, Buchanan J. Development and Validation of a Revised Nursing Home Assessment Tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

2. Lapid MI, Rummans TA. Evaluation and management of geriatric depression in primary care. Mayo Clin Proc. 2003;78(11):1423-9.

3. Barkin RL, Schwer WA, Barkin SJ. Recognition and management of depression in primary care: a focus on the elderly. A pharmacotherapeutic overview of the selection process among the traditional and new antidepressants. Am J Ther. 2000;7(3):205-26.

4. Reynolds CF, Dew MA, Pollock BG, Mulsant BH, Frank E, Miller MD, et al. Maintenance treatment of major depression in old age. N Engl J Med. 2006;354(11):1130-8.

5. Pinquart M, Duberstein PR, Lyness JM. Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. Am J Psychiatry. 2006;163(9):1493-1501.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*): This evidence has not been rated.

1c.6 Method for rating evidence:

1c.7 Summary of Controversy/Contradictory Evidence: No contradictory evidence has been identified.

1c.8 Citations for Evidence (*other than guidelines*): Saliba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

Teresi J, Abrams R, Holmes D, et al. Prevalence of depression and depression recognition in nursing homes. Soc Psychiatry Psychiatr Epidemiol. 2001;36(12):613-20.

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.

American Geriatrics Society and American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

symptoms associated with dementia. J Am Geriatr Soc. 2003;51(9):1287-98.

Greenberg PE, Kessler RC, Birnbaum HG, Leong SA, Lowe SW, Berglund PA, Corey-Lisle PK. The economic burden of depression in the United States: how did it change between 1990 and 2000? J Clin Psychiatr. 2003;64(12):1465-75.

Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: assessment and prevalence estimation. J Gerontol. 1989;44(1):M22-9.

Heston LL, Garrard J, Makris L, et al. Inadequate treatment of depressed nursing home elderly. J Am Geriatr Soc. 1992;40(11):1117-22.

Weintraub D, Datto CJ, Streim JE, et al. Second-generation issues in the management of depression in nursing homes. J Am Geriatr Soc. 2002;50(12):2100-1; author reply, 2101.

Datto CJ, Oslin DW, Streim JE, et al. Pharmacologic treatment of depression in nursing home residents: a mental health services perspective. J Geriatr Psychiatry Neurol. 2002;15(3):141-6.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Selected Consensus Statements taken from the American Geriatrics Society and American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral symptoms associated with dementia. J Am Geriatr Soc. 2003;51(9):1287-98.

2. Depression screening in nursing home residents should be accomplished using an additional instrument beyond the MDS [2.0].

6. Screening for depression in nursing home residents should be conducted:
a. In newly admitted residents after allowing 2 to 4 weeks to adjust to nursing home placement IV
b. In all residents at least every 6 months.

7. Diagnosis of depression should NOT be determined solely on the basis of severity score ratings of validated depression scales.

9. When a resident is identified as having depressive symptoms, physical, environmental, social, and spiritual issues should be evaluated.

10. Residents with new-onset depression or worsening of depressive symptoms should receive an evaluation focusing on:

a. Past history of depression symptoms and treatment
b. Current response to treatment of depression
c. Symptoms constituting a diagnosis of mood disorder
d. Suicidal ideation
e. Changes in cognitive function
f. Changes in social or family situation
g. New stressors or situational factors such as changes in staff
h. Availability of social and meaningful activities
i. Availability of positive (reinforcing) experiences
j. Unmet needs

11. Residents with new-onset depression or worsening of depressive symptoms should receive a medical evaluation by a physician, nurse practitioner, or physician assistant, including a history and physical examination that focuses on:

a. Assessment of pain
b. Nutritional status

c. Worsening of chronic medical conditions
d. Recent onset of new medical condition
e. Medications that have the potential to alter cognition or mood

12. Unless recent results are available, residents with new-onset depression or worsening of depressive symptoms should be considered for laboratory and diagnostic testing as determined by the findings of the history and physical examination:

a. Hemoglobin
b. Thyroid function
c. Electrolytes
d. Vitamin B12 level
e. Serum drug levels that may play a role in presentation of depression in this population
f. Complete blood cell count

21. Nonpharmacological interventions are effective in treating depressive symptoms in nursing home residents with major depression.

22. Pharmacological interventions are effective in treating depressive symptoms in nursing home residents with major depression.

1c.10 **Clinical Practice Guideline Citation:** American Geriatrics Society and American Association for Geriatric Psychiatry. Consensus statement on improving the quality of mental health care in U.S. nursing homes: management of depression and behavioral symptoms associated with dementia. J Am Geriatr Soc. 2003;51(9):1287-98.

1c.11 **National Guideline Clearinghouse or other URL:** This is not applicable.

1c.12 **Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):**
The guideline recommendations were based on a consensus process conducted by the American Geriatrics Society and American Association for Geriatric Psychiatry Expert Panel on Quality Mental Health Care in Nursing Homes. The evidence supporting each individual guideline recommendation was rated by a researcher on the panel and reviewed iteratively by the panel members. The ratings ranged from II-IV for the relevant recommendations.

1c.13 **Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
The guideline report indicates the rating system was developed by the panel researcher based on those used for other evidence-based consensus panels. No other details are provided.

1c.14 **Rationale for using this guideline over others:**
The National Guideline Clearinghouse guidelines for depression are not specific to persons in nursing facilities.

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y** ☐ **N** ☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	

S.1 Do you have a web page where current detailed measure specifications can be obtained?
S.2 If yes, provide web page URL:

2a. Precisely Specified

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):**
Using the PHQ-9 items in the MDS 3.0, for the Resident Interview Measure (Item D0200), the numerator is based on the total sum severity score (D0300) on the most recent MDS assessment in the selected quarter (which may be an annual, quarterly, significant change, or significant correction assessment). The total severity score reflects resident responses to questions asking about the frequency of nine symptoms over the last 2 weeks, including interest, mood, energy, appetite, self-value, ability to concentrate, change in responsiveness, or patience. The Staff Assessment Measure (Item D0500) is similar, except the judgment is being made by observers rather than the residents themselves. The numerator is calculated by using data from item D0300, the total self-reported depression severity score. While the self-report data are preferred, if data from D0300 are incomplete or unavailable then the numerator will be calculated using data from item D0600.

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
Numerator data are from the most recent MDS assessment which may be an annual, quarterly, significant change or significant correction assessment during the selected quarter (3-month period).

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Residents are counted if they are long-stay residents, defined as residents whose length of stay is greater than 100 days. Residents who return to the nursing home following a hospital discharge will not have their stay reset to zero. A resident can be eligible for inclusion in the numerator in one of two ways for the MDS 3.0, the Resident Mood Interview or Staff Assessment of Resident Mood. The score is ten for either the Resident Mood Interview or Staff assessment of Resident Mood. A total score is calculated from Column 2, Symptom Frequency. The Staff Assessment of mood (items D0500) should be used if a long-stay resident is missing data for three or more of the subitems of data elements D0200 for the Resident Assessment AND has valid data for seven or more of subitems A through I of item D0500 for the Staff Assessment, as described below. When the Resident Mood Interview is conducted, the resident must have score of two or greater for either D0200A or D0200B AND a score of two or more for five of the following items D0200A-I. When the Staff Assessment for Resident Mood is necessary, the resident must have score of two or greater for either D0200A or D0200B AND a score of two or more for five of the following items D0200A-I.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
The denominator is the total number of all long-stay residents in the nursing facility who have received an MDS assessment (which may be an annual, quarterly, significant change or significant correction assessment) during the selected quarter (3-month period) and who do not meet the exclusion criteria.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: The target population includes long-stay residents (those in the nursing facility for more than 100 days) of all ages.

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
Denominator data are from the MDS 3.0 annual, quarterly, significant change or significant correction assessment during the selected quarter (3-month period).

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Residents are counted if they are long-stay residents, defined as residents whose length of stay is greater than 100 days. Residents who return to the nursing home following a hospital discharge will not have their stay reset to zero. The target population for the denominator is the total number of all long-stay residents in the nursing facility who have received an MDS assessment (which may be an annual, quarterly, significant change or significant correction assessment) during the selected quarter (3-month period) and who do not meet the exclusion criteria.

2a-specs
C ☐
P ☐
M ☐
N ☐

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** A long-stay resident is excluded from the denominator if the MDS assessment is an admission assessment (OBRA) or a 5-day PPS scheduled assessment, if the resident is comatose, or if there are too many missing data in the relevant section of the MDS. Facilities are excluded from public reporting if they have fewer than 30 residents.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
A0310.A = 01 OBRA Admission assessment OR
A0310.B = 01 PPS Medicare Part A 5-day scheduled assessment OR
B0100 = 1 or missing (Comatose)
Additional exclusion for the Resident Interview Measure:
Three or more MDS 3.0 D0200 subitems are missing.
Additional exclusion for the Staff Assessment Measure:
Three or more MDS 3.0 D0200 subitems are missing,
AND three or more MDS 3.0 D0500 subitems are missing.

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
This is not applicable.

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
This is not applicable.

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Ratio
2a.20 Interpretation of Score:
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
For each facility, the number of long-stay residents meeting the numerator criteria and the number of long-stay residents meeting the denominator criteria are counted. The facility prevalence score is calculated as the number of long-stay residents in the facility during the selected quarter in the numerator divided by all long-stay residents during the selected quarter in the denominator (excluding residents for whom there are missing data).

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
Because the computed scores are not estimates, but include all residents who meet the measure criteria, in terms of discriminating performance, the computed scores can be used to make valid comparisons.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
This is not applicable.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)*
Electronic Clinical Data

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):**
Nursing Home Minimum Data Set 3.0

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
http://www.cms.hhs.gov/NursingHomeQualityInits/25_NHQIMDS30.asp#TopOfPage

2a.29-31 Data dictionary/code table web page URL or attachment: URL
http://www.cms.hhs.gov/NursingHomeQualityInits/25_NHQIMDS30.asp#TopOfPage

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested)*
Facility

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested)*
Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)
Behavioral Health: Mental Health

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: The reliability of the depression measure has been tested and showed it to be very reliable. Testing of the reliability of MDS 3.0 data items underlying the depression quality measure as well as a comparison with the MDS 2.0 quality measures/quality indicators was conducted by RAND as part of the MDS 3.0 development process.(1) A representative sample of for-profit and not-for-profit facilities and hospital-based and freestanding facilities was recruited for the study, which included 71 community nursing facilities in 8 states, 19 VA nursing homes, 3,258 nursing facility residents for the depression quality measure analysis, and 418 residents in the validation sample that compared PHQ-9 measures with other depression measures.

Saliba and Buchanan analyzed the reliability and validity of the PHQ-9 in a sample of 71 community and 19 VA nursing facilities distributed throughout the United States.(1) Residents were selected in these facilities to capture a representative sample of short- and long-stay residents, and in order to maximize the number of MDS 2.0 items assessed, the selection algorithms included a strong preference for capturing cases scheduled for MDS 2.0 admission assessments.

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
Saliba and Buchanan compared the PHQ-9 to two "gold-standard" measures: the Modified Schedule for Affective Disorders and Schizophrenia (m-SADS), and the Cornell Depression Scale, using kappas as measures of concordance.(1) These alternative measures were also administered to a sub-sample of residents, and the results compared with those from the PHQ-9.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
For the resident interview using the approach proposed for this quality indicator, Saliba and Buchanan found that 65% of their sample had no depression, 18% had minor depression, and 17% had major depression. They found that reliability was excellent: the average kappa between gold-standard nurses for the PHQ-9 resident interview was 0.935, and between gold-standard and facility nurses it was 0.968.(1) The proposed quality measure is a ratio constructed from these measures and is therefore reliable.

1. Saliba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

2b C ☐ P ☐ M ☐ N ☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Saliba and Buchanan analyzed the reliability and validity of the PHQ-9 in a sample of 71 community and 19 VA nursing facilities distributed throughout the United States.(1) Residents were selected in these facilities to capture a representative sample of short- and long-stay residents, and in order to maximize the number of MDS 2.0 items assessed, the selection algorithms included a strong preference for capturing cases scheduled for MDS 2.0 admission assessments.

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
Saliba and Buchanan compared the PHQ-9 to two "gold-standard" measures: the Modified Schedule for Affective Disorders and Schizophrenia (m-SADS), and the Cornell Depression Scale, using kappas as measures of concordance.(1) The kappa for the PHQ-9 and the m-SADS was very good. The correlation between the PHQ-9 and the Cornell depression scale was 0.63.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test*

2c C ☐ P ☐ M ☐ N ☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

conducted):
For the m-SADS the kappa was very good at 0.685, higher than the kappas comparing the m-SADS with the Geriatric Depression Scale, the MDS 2.0 quality indicator definition, and the MDS 2.0 Resource Utilization Groups (RUGs) definition. For the Cornell Depression Scale the kappa was very good at 0.63, also higher than for the Geriatric Depression Scale, the MDS 2.0 quality indicator definition, and the MDS 2.0 RUGs definition.

For the staff observational version of the PHQ-9, Saliba and Buchanan added a tenth element: recording patient irritability.(1) While this observational version also demonstrated good reliability and validity, it is not entirely comparable to the staff assessment measure that forms one basis of the quality indicator discussed here.

These analyses of validity and reliability are for the PHQ-9 portion of the MDS 3.0, which is a measure of prevalence, not for the proposed quality indicator that measures changes in depression from MDS review to MDS review.

Calculating the validity and reliability of the proposed quality measure is somewhat constrained by the fact that although the resident-interview and staff assessment versions are quite similar, there are differences. In addition to the obvious difference between obtaining responses directly from a resident versus a nurse's observations or abstracting from observations contained in the medical record, there are two additional important differences between the two measures. The resident self-assessment asks about "Feeling bad about yourself – or that you are a failure or have let yourself or your family down" and about "Thoughts that you would be better off dead, or of hurting yourself in some way." For these concepts the staff assessment asks about the resident "Indicating that s/he feels bad about self, is a failure, or has let self or family down" or stating "that life isn't worth living, wishes for death, or attempts to harm self." In other words, the staff assessment requires some external expression for these concepts to be present, while the self-assessment may elicit feelings/views that are not otherwise in the record and have not been previously expressed.

Other research has studied the ability of persons who may have cognitive impairments to self-report these measures, and on the ability of nursing facility staff to accurately report similar measures. Ruckdeschel et al. found that the MDS 2.0 Mood Disturbance items can be reliably and validly administered via self-report to persons scoring at least 12 on the Mini-Mental State Exam.(2) But Gross and Kazner found undercoding of the MDS 2.0 depression items among nursing facility staff who were unfamiliar with the patients, and who relied heavily on chart documentation.(3)

Given the differences between the resident interview and staff assessment versions and the potential for underreporting of the latter observational measure, it may be important to report depression separately for residents who can and cannot respond to the resident interview. This has implications regarding the ability to obtain sufficient sample sizes to obtain reliable estimates.

1. Saliba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.
2. Ruckdeschel K, Thompson R, Datto CJ, Streim JE, Katz IR. Using the minimum data set 2.0 mood disturbance items as a self-report screening instrument for depression in nursing home residents. Am J Geriatr Psychiatry. 2004;12:43-9.
3. Gross J, Kazmer J. Depression and the MDS. Provider. 2006;32:33-6.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
All long-stay residents with complete data on their OBRA assessments are included unless they meet one of the following criteria: the MDS assessment is an admission assessment (OBRA or a PPS 5-day scheduled assessment), or the resident is comatose. Admission assessment data are excluded because these items refer to the preceding two week window and may not represent an ongoing condition in the nursing facility. Comatose persons are excluded because the concept of depression is not applicable. Facilities with fewer than 30 residents are excluded from public reporting.

2d.2 Citations for Evidence:
This is not applicable.

2d
C ☐
P ☐
M ☐
N ☐
NA ☐

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

2d.3 Data/sample *(description of data/sample and size)*: This is not applicable.

2d.4 Analytic Method *(type analysis & rationale)*:
This is not applicable.

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
This is not applicable.

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: This is not applicable.

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
This is not applicable.

2e.3 Testing Results *(risk model performance metrics)*:
This is not applicable.

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: Because depression is a treatable condition, it is not appropriate to use risk adjustment to "condition" this measure on other factors.

2e C ☐ P ☐ M ☐ N ☐ NA ☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: Because the depression measures are not currently collected in the MDS 2.0, analyses have not been conducted that identify differences at the level of a nursing facility in the performance of the depression ratio measure based on the PHQ-9 in the MDS 3.0. However, in terms of the PHQ-9 for an individual, and using mental health professional validation interviews as the criterion standard, a PHQ-9 score > or =10 had a sensitivity of 88% and a specificity of 88% for major depression. Also, PHQ-9 scores of 5, 10, 15, and 20 represented mild, moderate, moderately severe, and severe depression, respectively. And a preliminary approach would be to consider a PHQ-9 score less than 10 and a 50% decline from the pretreatment score as clinically significant improvement.(1)

1. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001 Sep;16(9):606-13.

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Because the computed scores are not estimates, but include all residents who meet the measure criteria, in terms of discriminating performance, the computed scores can be used to make valid comparisons.

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Not applicable.

2f C ☐ P ☐ M ☐ N ☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: This is not applicable.

2g.2 Analytic Method *(type of analysis & rationale)*:
This is not applicable.

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
This is not applicable.

2g C ☐ P ☐ M ☐ N ☐ NA ☐

2h. Disparities in Care

2h

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)***:** This is not applicable. **2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** While MDS 3.0 collects data on the resident's race there are no current plans to stratify the measure by race because facilities tend to be homogenous by race, making disparities generally evident in the rating of the facility. In 2000, a study drawing on national MDS and OSCAR data found that two-thirds of all black residents were living in just 10% of all facilities.(1) A 2002 survey of a stratified sample of 39 nursing homes and 181 residential care/assisted living facilities in four states had similar findings.(2) 1. Smith D, Feng Z, Fennell M, Ainn J, Mor V. Separate and unequal: racial segregation and disparities in quality across U.S. nursing homes. Health Aff (Millwood). 2007;26(5):1448-558. 2. Howard D, Sloane P, Zimmerman S, Eckert J, Walsh J, Buie V, Taylor P, Koch G. Distribution of African Americans in residential care/assisted living and nursing homes: more evidence of racial disparity? Am J Public Health. 2002;92(8):1272-7.	**C** ☐ **P** ☐ **M** ☐ **N** ☐ **NA** ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C** ☐ **P** ☐ **M** ☐ **N** ☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information **3a.1 Current Use:** Testing not yet completed **3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** Nursing Home Compare http://www.medicare.gov/NHCompare/Include/DataSection/Questions/SearchCriteriaNEW.asp?version=default&browser=IE%7C6%7CWinXP&language=English&defaultstatus=0&pagelist=Home&CookiesEnabledStatus=True **3a.3 If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):** CMS expects that the Quality Measure will be used by nursing homes as a tool to decrease the prevalence of depressive symptoms in their residents. **Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4 Data/sample** *(description of data/sample and size)***:** Data were collected from 4,754 family members of nursing home residents Castle N. The Nursing Home Compare report card: consumers' use and understanding. J Aging Soc Policy.	**3a** **C** ☐ **P** ☐ **M** ☐ **N** ☐

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

2009;21(2), 187-208.

3a.5 Methods *(e.g., focus group, survey, QI project)*:
A comprehension index was used to examine whether family members understood the information contained in Nursing Home Compare for each current quality measure.

3a.6 Results *(qualitative and/or quantitative results and conclusions)*:
The study found that 31% of the consumers used the Internet in choosing a nursing home and 12% recalled using Nursing Home Compare; in general, the consumers' comprehension index scores were high, indicating good understanding. However, this quality measure was not specifically reported.

3b/3c. Relation to other NQF-endorsed measures

3b.1 NQF # and Title of similar or related measures:
This measure is intended to replace NQF#0197 Residents with worsening of a depressed or anxious mood, as the data source has changed; the MDS 2.0 is being replaced with the MDS 3.0. Other related measures are NQF # 0103 Major Depressive Disorder: Diagnostic Evaluation Status: Endorsed on: DEC 01, 2006 Steward(s): American Medical Association–Physician Consortium for Performance Improvement / Percentage of patients with a diagnosis of major depressive disorder who met the DSM-IV criteria during the visit in which the new diagnosis or recurrent episode was identified; NQF#0418 Screening for Clinical Depression; NQF#0518 Depression assessment conducted (home health).

(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**

3b. Harmonization
If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):
3b.2 Are the measure specifications harmonized? If not, why?
The proposed measure is based on the PHQ-9 that is part of the MDS 3.0 and is based on the diagnostic criteria for a major depressive disorder in the DSM-IV. But is not based on a recorded diagnosis of major depression disorder. The focus is nursing facility residents, not the outpatient or home health setting.

3b
C ☐ P ☐ M ☐ N ☐ NA ☐

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:
Compared to other endorsed measures cited in 3b.1 this measure focuses on long-stay nursing facility residents. The underlying MDS 3.0 items used to screen for depressive symptoms in the proposed measure have undergone extensive testing and are more reliable than the items used in the currently endorsed measure based on the MDS 2.0. In addition to using new items to identify possible depression, the proposed measure does not compare results between two time points, rather it is a targeted prevalence measure; this is in contrast to the related previously endorsed MDS 2.0 measure for nursing facility residents.

5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:

3c
C ☐ P ☐ M ☐ N ☐ NA ☐

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?

3

Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?**
Rationale:

3
C ☐ P ☐ M ☐ N ☐

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

Eval Rating

4a. Data Generated as a Byproduct of Care Processes

4a

4a.1-2 How are the data elements that are needed to compute measure scores generated?
Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)

C ☐ P ☐ M ☐ N ☐

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

4b. Electronic Sources

4b

4b.1 Are all the data elements available electronically? (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)
No

4b.2 If not, specify the near-term path to achieve electronic capture by most providers.
Not applicable.

C ☐ P ☐ M ☐ N ☐

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

4c. Exclusions

4c

4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?
No

4c.2 If yes, provide justification.

C ☐ P ☐ M ☐ N ☐ NA ☐

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

4d

4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.
No research could be identified that specifically addressed this issue.

C ☐ P ☐ M ☐ N ☐

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
The data collection method is already in operational use and there are no issues with these areas.

4e.2 Costs to implement the measure (*costs of data collection, fees associated with proprietary measures*)**:**
Data is collected as part of an existing process with no additional cost.

4e.3 Evidence for costs:
This is not applicable.

4e

C ☐ P ☐ M ☐ N ☐

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

4e.4 Business case documentation: The proposed measure relies on data from the MDS 3.0. As there is no change in the data collection method for the MDS 3.0 as compared with its predecessor, the MDS 2.0, we do not anticipate any additional burden to nursing facilities. MDS 2.0, and soon to be MDS 3.0, data are collected as part of an existing, federally mandated process used for payment and quality monitoring purposes.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** **C** ☐ **P** ☐ **M** ☐ **N** ☐
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited** ☐
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y** ☐ **N** ☐ **A** ☐

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Mail Stop S3-02-01, Baltimore, Maryland, 21244-1850

Co.2 Point of Contact
Cheryl, Wiseman, MS, MPH, cheryl.wiseman2@cms.hhs.gov, 410-786-1175-

Measure Developer If different from Measure Steward
Co.3 Organization
RTI International, 1440 Main Street, Suite 310, Waltham, Massachusetts, 02451-1623

Co.4 Point of Contact
Karen, Reilly, ScD, kreilly@rti.org, 781-434-1700-1791

Co.5 Submitter If different from Measure Steward POC
Karen, Reilly, ScD, kreilly@rti.org, 781-434-1700-1791, RTI International

Co.6 Additional organizations that sponsored/participated in measure development

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
See attached Table 1: Nursing Home Quality Measures Technical Expert Panel (January 2009).

This technical expert panel met over 2 days in January 2009 to review the environmental scan of the current quality measures and make recommendations regarding their transition from MDS 2.0 to MDS 3.0.

Ad.2 If adapted, provide name of original measure: This is not applicable. **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 2002 **Ad.7 Month and Year of most recent revision:** 02, 2010 **Ad.8 What is your frequency for review/update of this measure?** Every 3 years **Ad.9 When is the next scheduled review/update for this measure?** 02, 2013
Ad.10 Copyright statement/disclaimers:
Ad.11 -13 Additional Information web page URL or attachment: Attachment Depression tables_FINAL-634045020469642500.doc
Date of Submission (*MM/DD/YY*)**:** 05/05/2011