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.

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)

<table>
<thead>
<tr>
<th>Measure Descriptive Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>De.1</strong> Measure Title: Percent of Residents Who Self-Report Moderate to Severe Pain (Short-Stay)</td>
</tr>
<tr>
<td><strong>De.2</strong> Brief description of measure: This measure updates CMS’ current QM on pain severity for short-stay residents (people who are discharged within 100 days of admission). This updated measure is based on data from the Minimum Data Set (MDS 3.0) 14-day PPS assessments. This measure reports the percentage of short-stay residents with a 14-day PPS assessment during a selected quarter (3 months) who have reported almost constant or frequent pain and at least one episode of moderate to severe pain, or any severe or horrible pain, in the 5 days prior to the 14-day PPS assessment.</td>
</tr>
<tr>
<td><strong>De.3</strong> If included in a composite or paired with another measure, please identify composite or paired measure</td>
</tr>
<tr>
<td><strong>De.4</strong> National Priority Partners Priority Area: Care coordination</td>
</tr>
<tr>
<td><strong>De.5</strong> IOM Quality Domain: Patient-centered</td>
</tr>
<tr>
<td><strong>De.6</strong> Consumer Care Need:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions for Consideration by NQF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:</td>
</tr>
<tr>
<td><strong>A.</strong> 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.</td>
</tr>
<tr>
<td><strong>A.1</strong> 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</td>
</tr>
<tr>
<td><strong>A.2</strong> Indicate if Proprietary Measure (as defined in measure steward agreement):</td>
</tr>
<tr>
<td><strong>A.3</strong> Measure Steward Agreement: Government entity and in the public domain - no agreement necessary</td>
</tr>
<tr>
<td><strong>A.4</strong> Measure Steward Agreement attached:</td>
</tr>
</tbody>
</table>

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

C. The intended use of the measure includes both public reporting and quality improvement.

<table>
<thead>
<tr>
<th>Purpose: Public Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>D.1 Testing: No, testing will be completed within 12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? Yes</td>
</tr>
</tbody>
</table>

(for NQF staff use) Have all conditions for consideration been met?

<table>
<thead>
<tr>
<th>Met</th>
</tr>
</thead>
</table>

Staff Notes to Steward (if submission returned):

Staff Notes to Reviewers (issues or questions regarding any criteria):

Staff Reviewer Name(s):

<table>
<thead>
<tr>
<th>TAP/Workgroup Reviewer Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steering Committee Reviewer Name:</td>
</tr>
</tbody>
</table>

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

(for NQF staff use) Specific NPP goal:

1a.1 Demonstrated High Impact Aspect of Healthcare: Patient/societal consequences of poor quality

1a.2

1a.3 Summary of Evidence of High Impact: Research indicates that at least 40-85% of nursing facility residents have persistent pain. The percentage may be even higher; research suggests that pain is often not fully documented.(1, 2, 3, 4, 5, 6, 7)

Failure to identify the presence of pain or to assess its severity and functional impact can leave a potentially treatable symptom unrecognized and therefore unlikely to be addressed. Indeed, evidence suggests that pain is consistently under-treated, particularly among individuals with cognitive impairment.(3, 8, 9) A standard measure of resident pain is needed because of gaps in nursing staff’s knowledge of "best practice" pain management in hospitals and nursing facilities.(4, 10, 11, 12, 13) A standard measure also provides a benchmark for pain management practices that vary widely across nursing homes.(13, 14, 15)

Among the potential adverse physiological and psychological effects of unrelieved pain are impaired gastrointestinal and pulmonary function; nausea and dyspnea; increased metabolic rate, including increased tumor growth and metastasis in cancer; impaired immune response; insomnia, delayed healing, increased blood clotting, loss of appetite, and the inability to walk or move about; impairment of joint function with functional decline and increased dependency; and anxiety and depression.(16, 17, 18, 19) In the general population, unrelieved pain costs millions of dollars annually as a result of longer hospital stays, rehospitalizations, outpatient care, and emergency room visits.(20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31)

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).
Resident pain in nursing facilities is a subject of great interest to the public. Pain management in nursing facilities is central to the Omnibus Budget Reconciliation Act of 1987 (OBRA 87) mandate to promote “maximum practicable functioning” among residents, and failure to identify and address pain denies a resident the right granted in OBRA 87 to freedom from neglect.\(^{(32)}\) Advancing Excellence in America’s Nursing Homes has made the management of resident pain one of its major goals.\(^{(33)}\)

1a.4 Citations for Evidence of High Impact:  


1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Use of this measure should prompt nursing facilities to examine their attention to pain severity in recently admitted residents and lead to an increase in pain management efforts and reduction in pain severity.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
A version of this quality measure has been in use by CMS since 2002, drawing on data from an MDS 2.0 item

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable
based on staff assessment. A study of variability for this measure by the University of Colorado showed that in the first quarter (Q1) of 2006, the measure showed an acceptable degree of variability across facilities. See attached Table 1: Measure Variability Across Facilities.

Although the number of high-quality studies of pain management in nursing facilities is limited, those studies agree that resident pain is under-recognized and under-treated. A recent record audit of 291 residents in 14 long-term care facilities found a significant gap between evidence-based pain management recommendations and facility practices. Assessment was particularly weak; only 32% of the cases reported for pain once or twice a week, and only 3% of the cases reviewed had reported that pain impacted functioning and quality of life two or more times during the previous 30 days. One study focusing on pain in cancer patients reported underuse of analgesics and hospice, along with nursing facility staffing patterns as key issues in inadequate pain treatment for this population. Many studies and literature maintain that almost all pain, including pain at the end of life, can be managed with appropriate assessment and treatment, and research in pain management has identified the adoption of systematic implementation models, clinical decision-making algorithms, interdisciplinary approaches, and ongoing outcome evaluations as effective means to deliver effective pain relief in nursing homes.

1b.3 Citations for data on performance gap:

1b.4 Summary of Data on disparities by population group:
Although there is evidence of racial segregation between nursing facilities, with African-Americans tending to be concentrated in facilities with higher deficiency ratings, there has been little study of resulting potential disparities in reported pain. The research conducted on racial disparities in pain treatment has shown a greater incidence of untreated pain for black residents with cancer as compared to white residents with cancer.

Research has also identified disparities in pain management between cognitively intact residents and those who are cognitively impaired. In the current MDS 2.0 pain item, staff recording of cognitive status was inversely proportional to pain report; the most cognitively impaired residents were recording as suffering the
1c.4 the outcome of, or experience with, care. For non-outcome measures, briefly describe the relationship to desired outcomes, values and preferences of individuals/ the public.

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

1c.6 Method for rating evidence:


1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):
The specific recommendation is acute pain management in older adults.

1c.10 Clinical Practice Guideline Citation: University of Iowa Gerontological Nursing Interventions Research Center, Research Translation and Dissemination Core. 1997.
1c.11 National Guideline Clearinghouse or other URL:

1c.12 Rating of strength of recommendation (also provide narrative description of the rating and by whom):
The University of Iowa rated the relevant portions of the recommendations as follows: obtain self-report of pain from the older individual if possible-D; eleven point numeric rating scale-B; four point verbal rating scale-B.

1c.13 Method for rating strength of recommendation (If different from USPSTF system, also describe rating and how it relates to USPSTF):
The rating system used by the recommendation uses a five-point scale as follows:
A. There is evidence of well-designed meta-analysis in older adults.
B. There is evidence of well-designed controlled trials in the older adult population; randomized and nonrandomized, well-designed quasi-experimental and cohort studies in older adult populations with results that consistently support a specific action (e.g., assessment, intervention or treatment).
C. There is evidence of observational studies (e.g., correlational, descriptive studies) or controlled trials in older adults with inconsistent results.
D. There is evidence of integrative reviews, national clinical practice guidelines, or acute pain research in adults, but not specific to older adults.
E. There is evidence of expert opinion or multiple case reports regarding older adults.

The USPSTF grading system, described at http://www.ahrq.gov/clinic/3rduspstf/ratings.htm, grades the quality of the overall evidence for a service on a three-point scale (i.e., good, fair, or poor):
• Good: Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess the effects on health outcomes.
• Fair: Evidence is sufficient to determine the effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies, generalizability to routine practice, or indirect nature of the evidence on health outcomes.
• Poor: Evidence is insufficient to assess the effects on health outcomes because of the limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

1c.14 Rationale for using this guideline over others:
This guideline, registered with the National Guideline Clearinghouse, addresses acute pain management for rehabilitation.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Importance to Measure and Report?

Steering Committee: Was the threshold criterion, Importance to Measure and Report, met?

Rationale:

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)

2a. MEASURE SPECIFICATIONS

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):
The numerator is the number of short-stay residents who are able to self-report (item J200=1), who have a 14-day PPS assessment during the preceding 6 months, who report almost constant or frequent pain (item J0400 = 1 or 2) AND at least one episode of moderate to severe pain (item J0600A = 5, 6, 7, 8, or 9 on a scale of 1-10, with 10 being the worst pain you can imagine), OR item J0600B = 2 or 3 on a scale of 0-4, with 4 being very severe, horrible pain) OR very severe/horrible pain of any frequency (item J0600A = 10 on a scale of 1 to 10 OR item J0600B = 4 on a scale of 0 to 4) in the 5 days prior to the 14-day PPS assessment.

2a.2 Numerator Time Window (The time period in which cases are eligible for inclusion in the numerator):
The numerator data come from MDS 3.0 14-day PPS assessments conducted during the six months preceding each 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 short-stay residents, defined as residents whose length of stay is less than or equal to 100 days. The numerator details include the number of short-stay residents able to self-report (item J200=1) and who report almost constant or frequent pain on a scale of 1 to 4. These numeric ratings were defined as the following: 1 = the pain is almost constantly (item J0400 = 1 or 2) AND at least one episode of moderate to severe pain (item J0600A = 5, 6, 7, 8, or 9 on a scale of 1-10, with 10 being the worst pain you can imagine), OR item J0600B = 2 or 3 on a scale of 0-4, with 4 being very severe, horrible pain) OR very severe/horrible pain of any frequency (item J0600A = 10 on a scale of 1 to 10 OR item J0600B = 4 on a scale of 0 to 4) in the 5 days prior to the assessment.

2a.4 Denominator Statement (Brief, text description of the denominator - target population being measured):
The denominator is the total of all short-stay residents in the nursing facility who have received an MDS 3.0 14-day PPS assessment during the preceding 6 months from the selected quarter 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 short-stay residents of all ages who are admitted to the nursing facility.

2a.7 Denominator Time Window (The time period in which cases are eligible for inclusion in the denominator):
Denominator data come from MDS 3.0 14-day PPS assessments conducted during the 6 months preceding each 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 short-stay residents, defined as residents whose length of stay is less than or equal to 100 days. The target population includes all short-stay residents who have had a MDS 3.0 14-day PPS assessment (item A03100.B=2) during the 6 months preceding the selected quarter, except those who meet the exclusion criteria.

2a.9 Denominator Exclusions (Brief text description of exclusions from the target population): A resident is excluded from the denominator if they cannot self-report or there is missing data in the relevant questions in
Short-stay facilities with fewer than 20 residents are excluded from public reporting because of small sample size.

### 2a.10 Denominator Exclusion Details

(All information required to collect exclusions to the denominator, including all codes, logic, and definitions):

A 14-day PPS assessment was excluded if any of the following items had missing or inconsistent data for pain: J0400, J0600A, or J0600B. Item J0400 is the question about frequency of pain in the resident interview, with a 1 to 4 numeric rating response scale (with 1 being almost constantly). Item J0600A is the numeric rating question about intensity of pain in the resident interview, with a 0 to 10 numeric rating response scale (with 10 being the worst pain you can imagine). Item J0600B is the verbal descriptor scale question about intensity of pain in the resident interview, with a 1-4 verbal descriptor response scale. Data are inconsistent if the resident reports any frequency of pain in item J0400 while reporting a pain intensity of 0 in item J0600A or is unable to answer item J0600B (code 9). Data are also inconsistent if the resident is unable to answer item J0400 (code 9) while reporting a pain intensity of 1 or greater in item J0600A or any pain intensity in item J0600B.

### 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 short-stay residents meeting the numerator criteria and the number of (non-excluded) residents meeting the denominator criteria for this measure are counted. The facility-observed score for the measure is a prevalence score calculated as the number of residents in the facility in the numerator divided by all non-excluded residents in the denominator.

### 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.):

The data source or collection instrument is 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
2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested)

Post Acute/Long Term Care Facility: Rehabilitation

2a.38-41 Clinical Services (Healthcare services being measured, check all that apply)

<table>
<thead>
<tr>
<th>TESTING/ANALYSIS</th>
</tr>
</thead>
</table>

**2b. Reliability testing**

2b.1 Data/sample (description of data/sample and size): Yes, the testing is incomplete because the reliability testing for the measure is based on the MDS 2.0. The underlying pain items have significantly changed from the MDS 2.0 to the MDS 3.0 although RAND did perform item reliability testing on a national level as part of their MDS 3.0 development work. (1)

The proposed measure is based on two pain items in MDS 3.0, Section J items J0400 and J0600, with the numerator including all those residents who are able to self-report and who have been assessed during the selected quarter and who report almost constant or frequent pain (item J0400 = 1 or 2) AND at least one episode of moderate to severe pain (item J0600A = 5, 6, 7, 8, or 9 OR item J0600B = 2 or 3) OR very severe/horrible pain of any frequency (item J0600A = 10 OR item J0600B = 4) in the 5 days prior to the assessment.

Two major tests of the reliability of the current measure have been conducted. First, the MDS 2.0 measure items and the current quality measure were tested in the Data Assessment and Verification (DAVE 2) project conducted by Abt Associates. This project used a nationwide sample of randomly selected nursing facilities using MDS assessments for the period from April 1 to December 31, 2006. During this project, 173 two-stage reviews were performed. (2)

Second, the University of Colorado used national facility-level quality measure data from the third quarter (Q3) of 2003 through Q3 of 2006, which came from the Quality Improvement and Evaluation System (QIES) MDS Express Reports on the CMS Intranet; and Online Survey, Certification, and Reporting (OSCAR) data related to facility characteristics (e.g., state, resident census, number of beds, staffing) and certification survey results were downloaded from the QIES Workbench. (3) A 10% random sample of all Medicare-certified nursing facilities was downloaded from MDS assessment records. Analyses were based on complete MDS data from January 2005 through March 2006, nearly complete data for April 2006, and partial data for May and June 2006.


2b.2 Analytic Method (type of reliability & rationale, method for testing):

The DAVE 2 project used a two-stage cluster sample design to examine MDS reporting. A trained nurse reviewer selected a current resident with a recent assessment performed by the nursing facility within the past 14 days. In Stage 1 of this review, the nurse reviewer conducted a blind reassessment of the resident using standard MDS assessment and coding procedures (examination of the medical record; observation of the resident; interview of staff, resident, and family; and use of coding criteria). In Stage 2 of this assessment, the DAVE 2 nurse reviewer's assessment was compared to the corresponding nursing facility assessment and each discrepancy was reconciled, with the nursing home assessor and the nurse reviewer agreeing on the appropriate response. In addition to data entering the facility MDS code, the DAVE 2 code, and the reconciled code into the MDS-QC data entry software, the DAVE 2 nurse reviewer entered a “reason code” to attribute the cause of the discrepancy, per MDS item reviewed, to an established list of reasons.

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.
The national test of MDS 3.0 items by Saliba and Buchanan examined the agreement between assessors (reliability); the response rates for interview items; user satisfaction and feedback on changes; and the time to complete the assessment. The network of Quality Improvement Organizations was used to identify the gold-standard (research) nurses and recruit community nursing facilities to participate in the national evaluation, including a representative sample of for-profit and not-for-profit facilities and hospital-based and free-standing facilities. The gold-standard nurses were trained in the MDS 3.0 instrument, and they, in turn, trained a facility nurse from each participating nursing facility in their home states. Residents participating in the test were selected to capture a representative sample of short- and long-stay residents.

2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted):

The DAVE 2 project found a two-stage discrepancy rate of 7.3% for the MDS 2.0 pain frequency item (J0400) and 9.1% for the MDS 2.0 pain intensity item (J0600). (1) These MDS 2.0 measure items correspond to item J0400 and item J0600 of MDS 3.0, which are essentially the same in scope, although they rely on a nurse assessment rather than a resident report.

The national pilot test of the MDS 3.0 items showed good reliability with little evidence of confusion. For the pain items, the average kappa for gold-standard nurse to gold-standard nurse agreement was .961, and the average kappa for gold-standard nurse to facility nurse agreement was .967. (2)


2c. Validity testing

2c.1 Data/sample (description of data/sample and size): The data came from two sources: (1) national facility-level quality measure data from Q3 of 2003 through Q3 of 2006, which came from the QIES MDS Express Reports on the CMS Intranet; and (2) OSCAR data related to facility characteristics (e.g., state, resident census, number of beds, staffing) and certification survey results were downloaded from the QIES Workbench. A 10% random sample of all Medicare-certified nursing facilities was also downloaded from MDS assessment records. Analyses were based on complete MDS data from January 2005 through March 2006, nearly complete data for April 2006, and partial data for May and June 2006.

Information for this response and the other responses in regard to Validity Testing is from:


2c.2 Analytic Method (type of validity & rationale, method for testing):

The analysis of the current measure evaluated measure validity in a number of ways to examine the expected positive influence of public reporting on quality of care, which is an assessment of the degree to which quality measure triggering rates have improved over time; evaluate convergent validity, which is an assessment of the correlation of the quality measure with all other measures; determine if the quality measure triggering rate was influenced by factors that are unrelated to facility quality, which is an evaluation of seasonal variations in triggering rates across the 13 quarters of data. The analysis also computed descriptive statistics and conducted a one-way analysis of variance (ANOVA) for the measure to examine the amount of variance in triggering rates explained by the state where a facility was located.
2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted):
Yes, the testing is incomplete because the validity testing is based on the MDS 2.0. When the MDS 3.0 data are analyzed after implementation October, 2010, RTI will further test the quality measure validity.

These results reflect the performance of the current post-acute pain measure and the underlying MDS 2.0 items for those measures, which measure the same pain factors as the MDS 3.0 items for the proposed measure. In the proposed measure, data will be collected directly from the resident.

Only 8.0% of the variance in report rate for the current measure was explained by the state where a facility was located. The analysis found that public reporting may have had some influence on the decreased level of reported pain over time due to the decline in the triggering rate.

See attached Table 2: Measure Trends Over Time.

The current post-acute care pain measure demonstrated a .55 correlation with the current chronic care pain measure (also based on MDS 2.0); although correlations with other clinical measures are weak.

See attached Table 3: Correlations of Quality Measures.

There is little evidence of seasonal variations, as shown by the previously mentioned triggering rates, and the analysis found that only 8% of the variance in report rate for this measure was explained by the state where a facility was located. The limited correlation to other clinical measures may reflect the multiplicity of causes and potential treatments for pain, and the limited variation in seasonal rate and rate among states makes this measure a reliable guide to the level of reported pain.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Excluding missing data for existing quality measures is standard practice and was initially endorsed by NQF. Missing data is excluded from the calculation of the quality measures for several reasons. 1) There are legitimate reasons for facility staff not to select a ‘dash’ rather than a response; for example, if a resident is discharged or transferred abruptly, the staff may not be able to complete all items, however, an assessment is required for payment. The intent of the ‘dash’ is to allow the facility to submit an assessment when the staff are unable to complete the entire assessment. 2) Historically there has been very little missing data. For example, the current quality measure “Percent of residents who were physically restrained”, is based on three fields on the MDS 3.0. For all of the non-admission target assessments for calendar year 2009, there were 5,242,022 such assessments and 629 assessments (0.012%) had a dash for one or more of the three fields for the physical restraint measure. 3) We remain concerned about a change in measure definition that may result in incentivizing the facility staff to fill in a response to avoid a missing item. We believe that the result will lead to decreased validity and usefulness of the measure.

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

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):

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=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: The short-stay population is admitted from an acute facility and represents a different case mix compared to long-stay residents. The short-stay population, particularly the post-surgical population, are likely to have acute pain which can be effectively treated and which should be measured independent of these risk factors. However, when the MDS 3.0 data are analyzed after implementation October, 2010, potential opportunities for risk-adjustment can be further analyzed.

No adequate risk adjustment has been developed. Efforts to develop adequate risk adjustment are described in the following publication:


2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use (description of data/sample and size): These results reflect the performance of the current chronic care pain measure and the underlying MDS 2.0 items for that measure, which measures the same pain factors as the MDS 3.0 items for the proposed measure. In this proposed measure, data will be collected directly from the resident.

The data came from two sources: (1) national facility-level quality measure data from Q3 of 2003 through Q3 of 2006, which came from the QIES MDS Express Reports on the CMS Intranet; (2) and OSCAR data related to facility characteristics (e.g., state, resident census, number of beds, staffing) and certification survey results were downloaded from the QIES Workbench. A 10% random sample of all Medicare-certified nursing facilities was also downloaded from MDS assessment records. Analyses were based on complete MDS data from January 2005 through March 2006, as nearly complete data for April 2006, and partial data for May and June 2006.

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):
An analytical team at the University of Colorado’s Health Sciences Center examined the triggering rates for the measure at the facility level. Below are the measure scores from testing or current use (description of scores [e.g., distribution by quartile, mean, median, standard deviation], identification of statistically significant and meaningfully differences in performance). For 10,976 facilities, the mean triggering rate was 21.7%, with a standard deviation of 14.2%. The following table reports the full results of the analysis:

See attached Table 1: Measure Variability Across Facilities.

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.

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.
2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results (scores by stratified categories/cohorts): The measure is not stratified.

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:

Although 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.(1, 2 3)

Research has also identified disparities in pain management between cognitively intact residents and those who are cognitively impaired. In the current MDS pain item, staff recording of cognitive status was inversely proportional to pain report; the most cognitively impaired residents were recorded as suffering the least pain and received the least pain therapy.(4) In the MDS 3.0, new pain items were included that focus on patient interview and have been shown to be able to be answered by cognitively impaired residents.(5) However, the sample size at the facility level may not support stratification, but this will be evaluated in the future as MDS 3.0 data become available.


TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Scientific Acceptability of Measure Properties?

Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met?

<table>
<thead>
<tr>
<th>2h.1</th>
<th>2h.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

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)

3a. Meaningful, Understandable, and Useful Information

3a.1 Current Use: Not in use but testing 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):

The predecessor version of this measure is currently used in Nursing Home Compare, and this measure is designed to replace it there.


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 facilities as a tool to monitor and reduce resident pain. The national level of pain reported by the current measure has declined from 22.6% in Q1 of 2005 to 19.6% in Q3 of 2009. (Data are available at http://www.cms.hhs.gov/MDSPubQIandResRep/02_qmreport.asp#TopOfPage)

This measure is also cited by the Mission of the Advancing Excellence in America’s Nursing Homes Campaign, a cooperative quality program sponsored by long-term care providers; consumers and advocates; and nursing facility practitioners, including nurses, health care professionals, medical directors, nursing facility administrators, government agencies, quality improvement organizations, and private organizations supporting nursing facility education. Based on projection from MDS Quality Measure reporting data, the Advancing Excellence in America’s Nursing Homes Campaign set several goals to reduce the national level of reported pain in long-term care by September 2008. Unfortunately, the results to date demonstrate that, by the second quarter of 2009, none of the goals had been achieved: the national average of reported pain in short-term care remained above 15%, fewer than 10% of nursing facilities reported rates of short-stay residents with pain below 10%, many nursing facilities still reported rates of short-term residents with pain exceeding 46%, and the numbers of short-stay nursing facility residents with pain rose slightly to more than 150,000 rather than declining by 130,000.

http://www.nhqualitycampaign.org/star_index.aspx?controls=campaignReports

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): A recent study found that consumers could accurately interpret the quality information given for all the measures reported by Nursing Home Compare.

Data were collected from 4,754 family members of nursing facility residents


3a.5 Methods (e.g., focus group, survey, QI project):

A comprehension index was used to examine whether the information contained in Nursing Home Compare for each quality measure was understood by family members.

3a.6 Results (qualitative and/or quantitative results and conclusions):

The study found that 31% of the consumers used the Internet to help them choose a nursing facility, 12% recalled using Nursing Home Compare, and, in general, the consumers’ comprehension index scores were high, indicating a good understanding. The comprehension index for the current post acute care pain measure was among the highest, 5.62 on a scale of 1 to 8.

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

3b.1 NQF # and Title of similar or related measures:

The proposed measure is intended to replace NQF #0186—Recently hospitalized residents who experienced moderate to severe pain at any time during the 7-day assessment which is based on MDS 2.0. The proposed measure is based on MDS 3.0. Other related measures are: NQF #0192-Residents who experience moderate to
severe pain during the 7-day assessment period (risk-adjusted); NQF #0177-Improvement in pain interfering with activity; NQF #0523-Pain Assessment Conducted; NQF #0420-Pain Assessment Prior to Initiation of Patient Therapy; NQF #0524-Pain Interventions Implemented.

(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?
No. All the above measures are based on other instruments except for NQF #0186 and NQF #0192, which are based on a previous version of the MDS, version 2.0. NQF #0186 is scheduled to be replaced by this proposed measure and NQF #0192 is scheduled to be replaced by a measure being proposed at the same time.

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:
The data source for the proposed measure is changing to the MDS 3.0 and is based on pain assessment items found to have greater reliability than the pain assessment items found in the MDS 2.0.

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:

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

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

Rationale:

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

4a. Data Generated as a Byproduct of Care Processes
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), Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)

4b. Electronic Sources
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)

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).

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.)

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

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

4c. Exclusions

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.

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

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.

The proposed measure excludes those residents who cannot self-report their pain. Analysis of the underlying pain items to date indicates that this is a relatively small percentage of the resident population.(9) Thus, it is not expected that the exclusion of residents who cannot self-report will introduce a significant error in the aggregate measurement. However, analysis of the measure and underlying items will take place after the MDS 3.0 is implemented in October 2010 to confirm this preliminary finding and identify any other patterns of inaccuracy, error, or unintended consequences.

The proposed MDS 3.0 measure, which relies on resident report, is designed to replace a current MDS 2.0 measure, which was based on staff assessment. The current measure reported consistently and sometimes dramatically lower rates than those found in nursing homes in randomized controlled trial studies involving self-reporting. The proposed measure may itself underreport pain because it excludes those nursing home residents who are unable to report their pain, generally due to dementia. However, patient self-report of the presence and severity of pain, which is incorporated in the MDS 3.0 items supporting the proposed measure, is considered the most reliable and accurate approach to pain assessment. Both the American Geriatrics Society Panel on Persistent Pain in Older Persons and the Department of Veterans Affairs endorse this approach.(1, 2)

A growing number of studies and other literature demonstrate that even nursing home residents with moderate to severe cognitive impairment can reliably respond to questions about pain.(3, 4, 5, 6, 7, 8) Several studies in elders with varying cognitive status suggest that some tools may be more reliable and “user friendly” than others for obtaining self-reports of pain from this population, and the new items in MDS 3.0 incorporate these more reliable and user-friendly approaches.(9, 10, 11, 12, 13, 14,) A national test of the MDS 3.0 items supporting the proposed measure found that 87% of the test sample of residents and 89% of a validation sample of residents were able to successfully complete the pain interview portion of the MDS 3.0 upon which this measure is based.(9) Further testing is needed though because at least one expert, Vincent Mor, believes that the number of residents who cannot be interviewed will be higher when MDS 3.0 is placed into general use.(15)

Recent research has found a general decline in the percentage of residents with pain (as defined by this measure) admitted to nursing facilities for long-term care by approximately 13% after the first publication of the current pain measure in 2002. Analysis associated with this study suggests that nursing facilities exhibited a tendency to avoid such residents to improve their rating for the measure, although the authors concede that, due to the difficulty in accurately measuring pain, it is possible that the decline was due to ascertainment bias.(16)

The proposed measure addresses an additional significant issue with the current measure, in which pain is reported by the staff assessor, relying on the assessor’s own observations and those of other staff and without the use of a standard scale, and subject to ascertainment bias. The proposed measure employs a resident interview with a standardized scale of 1 (almost constantly) to 4 (rarely) for frequency of pain and a choice of standardized scales of 0 (no pain) to 10 (worst pain you can imagine) or 1 (mild) to 5 (very severe, horrible) for...
An example of an unintended consequence of this measure may occur if residents report that pain frequency decreased, however, pain intensity increased; or the reverse occurs, if pain intensity decreased but pain frequency increased. As part of the validation testing for this measure, RTI will examine responses for change, lack of change, and direction of change as well as patterns of both the frequency and intensity to assess whether there is an effect on the face validity of the measure.


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 no issues are anticipated.

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

4e.3 Evidence for costs:
This is 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?

| Steering Committee: Overall, to what extent was the criterion, Feasibility, met? |
|--------------------------|------------------|
| Rationale:              |                  |

RECOMMENDATION

(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.

<table>
<thead>
<tr>
<th>Steering Committee: Do you recommend for endorsement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comments:</td>
</tr>
</tbody>
</table>

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 300, 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 4: Nursing Home Quality Measures Technical Expert Panel (January 2009) for a list of workgroup or panel member names and organizations.

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

#### If adapted, provide name of original measure:

This measure was adapted from the measure of the same name derived from MDS 2.0 data.

**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 Moderate to Severe Pain short stay tables_FINAL.doc](attachment: Moderate to Severe Pain short stay tables_FINAL.doc)

**Date of Submission (MM/DD/YY):** **05/05/2011**
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:
  - **Intermediate outcome** - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - **Process** - 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).
  - **Structure** - 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.
  - **Patient 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.
  - **Access** - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - **Efficiency** - 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.