# 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 #:** NH-026-10 **NQF Project:** Nursing Homes 2010 |
| **MEASURE DESCRIPTIVE INFORMATION** |
| De.1 **Measure Title:** Consumer Assessment of Health Providers and Systems (CAHPS®) Nursing Home Survey: Discharged Resident Instrument |
| De.2 **Brief description of measure***:* The CAHPS® Nursing Home Survey: Discharged Resident Instrument is a mail survey instrument to gather information on the experience of short stay (5 to 100 days) residents recently discharged from nursing homes. This survey can be used in conjunction with the CAHPS Nursing Home Survey: Family Member Instrument and the Long Stay Resident Instrument. The survey instrument provides nursing home level scores on 4 global items. In addition, the survey provides nursing home level scores on summary measures valued by consumers; these summary measures or composites are currently being analyzed. The composites may include those valued by long stay residents: (1) Environment; (2) Care; (3) Communication & Respect; (4) Autonomy and (5) Activities. |
| 1.1-2 **Type of Measure:** Patient experience<br>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:** Patient and family engagement<br>De.5 **IOM Quality Domain**: Patient-centered<br>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.*<br>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**<br>Y☐<br>N☐ |

| | |
|---|---|
| **A.2 Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:**<br>**A.3 Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary<br>**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**<br>Y☐<br>N☐ |
| **C.** The intended use of the measure includes both public reporting and quality improvement.<br>►**Purpose:** Public reporting, Internal quality improvement | **C**<br>Y☐<br>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.<br>D.1**Testing:** No, testing will be completed within 24 months<br>D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?**<br>Yes | **D**<br>Y☐<br>N☐ |
| (for NQF staff use**)** **Have all conditions for consideration been met?**<br>**Staff Notes to Steward** (*if submission returned*)**:** | **Met**<br>Y☐<br>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)<br>**1a. High Impact** | Eval Rating |
| (for NQF staff use) Specific NPP goal: | |
| 1a.1 **Demonstrated High Impact Aspect of Healthcare:** Patient/societal consequences of poor quality<br>**1a.2**<br><br>1a.3 **Summary of Evidence of High Impact:** CMS estimates that in 2007 there were 1.96 million discharges who used the Medicare SNF benefit, the predominant means of financing short stays in nursing homes for post acute or rehabilitation purposes. This SNF population accounted for 67.9 million covered days of care with an average of 35 days per discharge and average reimbursement of $11,305, or a average total reimbursement of $22 billion dollars. The National Health Expenditures Accounts (CMS, 2009) estimate that nursing home costs totaled $131 billion in 2008.<br><br>With the passage of the Omnibus Reconciliation Act of 1987 (OBRA'87) Congress responded to growing concerns about the quality of care that nursing home residents received by requiring reforms in the federal certification and oversight of nursing homes. OBRA'87 shifted evaluations of health care quality from a focus on structure, and process criteria to clinical outcomes, resident satisfaction and quality of life. Since OBRA'87 implementation, GAO (2005; 2007)has continued to investigate quality of care in nursing homes and quality oversight activities of CMS and the states. | **1a**<br>C☐<br>P☐<br>M☐<br>N☐ |

**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).

Concurrent with changes from OBRA'87 implementation, a radical rethinking of the long term care system known as "culture change" began more than a decade ago. Culture change refers to the transformation of nursing homes from an "acute care" model to a consumer-directed model. Common themes of changes include: autonomy in personal choices for the residents, improved communication between residents and staff,and more homelike environments (www.pioneernetwork.net). The Pioneer Network estimates that 5% of nursing homes have fully adopted culture change (www.pioneernetwork.net). Resident/Patient Experience surveys are one tool for a nursing home to use to become more resident-centered. The Institute of Medicine (2010) includes patient-centeredness in its conceptual framework for categorizing health care quality and disparities measurement. The National Priorities Partnership (http://www.nationalprioritiespartnership.org/PriorityDetails.aspx?i

The CMS Nursing Home Compare web site publishes separate quality performance measures for the short stay and long stay populations.

1a.4 **Citations for Evidence of High Impact:** for CMS estimates of Medicare SNF users: http://www.cms.hhs.gov/MedicareFeeforSvcPartsAB/Downloads/NationalSum2007.pdf
CMS, Nursing Home Data Compendium, 2008 edition.

CMS National Health Expenditure Data is at http://www.cms.gov/NationalHealthExpendData/

GAO (Dec. 2005). "Despite increased oversight, challenges remainin ensuring high-quality care and resident safety" www.gao.gov/cgi-bin/getrpt?GAO-06-117.

GAO (May 2007). "Continued attention is needed to improve quality of care in small but significant share of homes." www.gao.gov/cgi-bin/getrpt?GAO-07-794T.

Institute of Medicine Committee on Future Directions for the National Healthcare Quality and Disparities Reports; Cheryl Ulmer, Michelle Bruno, and Sheila Burke, Editors; Future Directions for the National Healthcare Quality and Disparities Reports. Washington, DC: National Academy Press, 2010.

**1b. Opportunity for Improvement**

1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** The goal would be to use this resident survey as feedback to transform nursing home care to be resident-directed/centered and achieve the highest quality of life and quality of care for this nursing home population.

1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**
The 2008 National Ombudsmen Reporting System (NORS) data showed that the top complaint of nursing home residents and their families, eliciting some 14,329 complaints to ombudsmen, was failing to respond to requests for assistance. Specific complaints relating to these items include lack of assistance with toileting which had 3,404 complaints; lack of assistance with drinking which had 2,899 complaints; and lack of assistance with eating which had 1,529 complaints (NORS, 2008). Complaints relating to dignity, respect and staff attitudes were also among the top ten.

Under contract with CMS, states conduct nursing home inspections, known as surveys, to assess compliance with federal quality and safety requirements, including requriements for resident rights and quality of life. According to the CMS Nursing Home Compare website, the US average number of nursing home deficiencies issued as of March 2010 was 8; however the range of deficiencies by state was 0 to 68.

1b.3 **Citations for data on performance gap:**
1. National Ombudsmen Reporting System (NORS,2008). Top 20 complaints by category for nursing facilities (FFY 1996-2008). 2008 National Ombudsman Reporting System Data Tables (Unlettered Tables in Appendix B). Retrieved on December 31, 2009 from http://www.aoa.gov/AoARoot/AoA_Programs/Elder_Rights/Ombudsman/National_State_Data/2008/Index.aspx.

2. CMS Nursing Home Compare website contains information on U.S. average number of deficiency citations at www.medicare.gov/NHCompare

**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.

1b.4 **Summary of Data on disparities by population group:**
not available

1b.5 **Citations for data on Disparities:**
not available

**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*)**:** For consumer satisfaction/experience data to be useful to nursing homes (i.e., know what areas need improvement and which have priority), surveys should measure what is important to residents. Survey data could also be used by consumers to help select higher quality nursing homes.

Some research indicates that higher resident satisfaction is associated with better resident clinical outcomes.

1c.2-3. **Type of Evidence:** Expert opinion, 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*)**:**
Carefully developed patient experience surveys can inform nursing home providers about areas that need improvement particularly in areas that residents and families consider important.(see section 3a.6 for focus group results on what is important to consumers). These survey items complement the data nursing homes may currently collect to support improvements in internal customer services and quality related activities.

Two separate unpublished studies by Castle (personal communication, April 2010) indicate that higher resident satisfaction is associated with fewer nursing home deficiency citations and clinical outcomes (less restraints and less depression).

1c.5 **Rating of strength/quality of evidence** (*also provide narrative description of the rating and by whom*)**:**
ungraded

1c.6 **Method for rating evidence:** ungraded

1c.7 **Summary of Controversy/Contradictory Evidence:** none identified

1c.8 **Citations for Evidence (***other than guidelines***):** Nicholas Castle, Ph.D., University of Pittsburgh (personal communication, April 2010), unpublished research from 2 study samples. (1) a sample of 3000 residents in 200 nursing homes; and (2) a sample of 180 nursing homes with family, resident, and staff satisfaction surveys.

1c.9 **Quote the Specific guideline recommendation (***including guideline number and/or page number***):**
not applicable

1c.10 **Clinical Practice Guideline Citation:** not applicable
1c.11 **National Guideline Clearinghouse or other URL:** not applicable

1c.12 **Rating of strength of recommendation** (*also provide narrative description of the rating and by whom*)**:**
not applicable

1c.13 **Method for rating strength of recommendation** (*If different from* USPSTF system, *also describe rating and how it relates to USPSTF*)**:**
not applicable

1c.14 **Rationale for using this guideline over others:**

**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:
o<u>Intermediate outcome</u> – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
o<u>Process</u> – 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 t ... [1]

**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 ... [2]

**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.

**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 ... [3]

| | |
|---|---|
| not applicable | |
| **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?**<br>**Rationale:** | **1**<br>**Y**☐<br>**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?**<br>S.2 **If yes, provide web page URL:**<br><br>**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***):**<br>The following topics are measured for nursing homes from a resident´s perspective:<br>Global Items:<br>Global Rating of care received from staff: sum of resident scores on 0 to 10 scale<br>Global Rating of special therapy care: sum of resident scores on 0 to 10 scale<br>Global Rating of overall nursing home: sum of resident scores on 0 to 10 scale<br>Global item whether respondent would recommend nursing home: sum of resident scores on item (see codebook for points assigned to each response category)<br><br>Composites: We expect some composites to be similar to the long stay resident instrument such as Environment, Care, and Communication & Respect. We are not sure if the Autonomy and Activities Composites will be relevant to short stay residents. Data analysis is currently being conducted.<br><br>2a.2 **Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):**<br>when resident was in nursing home<br><br>2a.3 **Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**<br>to be finalized for each composite and global item when analysis is completed | |
| 2a.4 **Denominator Statement (***Brief, text description of the denominator - target population being measured***):**<br>The denominator is the total number of surveys for respondents that meet CAHPS completion standard (50% of key items answered) and any applicable screener.<br><br>2a.5 **Target population gender:** Female, Male<br>2a.6 **Target population age range:** 18+<br><br>2a.7 **Denominator Time Window (***The time period in which cases are eligible for inclusion in the denominator***):**<br>when resident was in nursing home<br><br>2a.8 **Denominator Details (***All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**<br>to be finalized for each composite and global item when analysis is completed | **2a-specs**<br>**C**☐<br>**P**☐<br>**M**☐<br>**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*)**:** We will exclude all residents whose length of stay (LOS) in the facility is less than 5 or greater than 100 days from the date of admission. Residents who are discharged to any hospital with return anticipated will not have their day count reset to zero when they return to the facility (AHRQ will harmonize its specification on short stay residents with CMS). AHRQ created a separate survey for long stay residents whose opinion can be obtained best through in-person administration because, on average, they have more cognitive impairment than short stay residents. In addition, the CAHPS team believed that a minimum number of days (5) was needed for a short stay resident to have sufficient experience with facility care. We also exclude residents who are under age 18 and those who were discharged more than 2 months prior to sample frame development date. Those who were discharged to another care facility and not discharged home and those who were deceased were also excluded.

**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.10 Denominator Exclusion Details** (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*)**:**
self-explanatory

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

**2a.12-13 Risk Adjustment Type:** to be finalized when analysis of 2009 data is completed

**2a.14 Risk Adjustment Methodology/Variables** (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method*)**:**
We will use a similar methodology to that used for the Family Member survey found on pages 26-33 of the AIR Final Report. Variables to be used as case mix adjusters will be finalized when analysis is completed.

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

**2a.18-19 Type of Score:** Non-weighted score/composite/scale
**2a.20 Interpretation of Score:**
**2a.21 Calculation Algorithm** (*Describe the calculation of the measure as a flowchart or series of steps*)**:**
to be finalized for each composite and global item when analysis of 2009 data is completed

**2a.22 Describe the method for discriminating performance** (*e.g., significance testing*)**:**
For statistical significance for each composite or global item, we will use a t-test comparing each nursing home mean to the mean of all the nursing home means.

**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)*:
Survey Administration Guidelines are being developed by the CAHPS User Support Contract (similar to those on web site for Family Member Instrument and Long Stay Resident Instrument) as part of the finalization of the Discharged Resident Instrument.

The recommended minimum sample size will be finalized when analyses of the 2009 MHCC field test are completed. The 2006 Harvard Final report, Table 28c on page 88, suggests minimum sample sizes for the 5 composites for the mail sample (discharged resident) survey based on an N=123. The sample sizes needed are no more than 39 completes to achieve reliability of 0.7 for all 5 composites, much smaller than the sample size needed for the Long Stay Resident (interview sample).

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

**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 CAHPS - Discharged Resident Survey BETA Version February 2007

**2a.26-28 Data source/data collection instrument reference web page URL or attachment:** Attachment nhcahps-discharged-beta2007.pdf

**2a.29-31 Data dictionary/code table web page URL or attachment:**

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

2a.36-37 **Care Settings (***Check the setting(s) for which the measure is specified and tested)*
Nursing home (NH) /Skilled Nursing Facility (SNF)

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Clinicians: Pharmacist, Other, Clinicians: PA/NP/Advanced Practice Nurse nurse aides

## TESTING/ANALYSIS

**2b. Reliability testing**

2b.1 **Data/sample** *(description of data/sample and size)***:** 2005 field test data from 12 nursing homes in four New England states (n=127) and 2009 field test data from 60 nursing homes in Maryland (n=1828)

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:**
To look at reliability, internal consistency reliability (alpha) was estimated. This is a measure of how well the items in a composite hang together. Composites should have an alpha of 0.70 or greater to be considered reliable. Additionally, we looked at nursing-home (NH)-level reliability, or inter-unit reliability (IUR). This statistic represents a transformation of the F-statistic for testing differences among agencies on an item or composite (IUR = (F-1)/F). IUR can be interpreted as the fraction of the variation among facility scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or composite to discriminate across facilities is greater. An IUR > 0.70 is considered to indicate a high level of discriminant ability for an item or composite. As the IUR gets smaller, you need a larger sample in order to reliably discriminate across facilities.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
For 2005 data, Cronbach's alpha is reported for the 5 Composites in Table 25 (Mail) on pages 82-83 of Harvard Final report. The IUR is reported in Table 28c on page 88 of the Harvard Final Report.

For 2009 data, analyses are currently being conducted

**2b**
C☐
P☐
M☐
N☐

**2c. Validity testing**

2c.1 **Data/sample** *(description of data/sample and size)***:** 2005 field test data from 12 nursing homes in four New England states (n=127) and 2009 field test data from 60 nursing homes in Maryland (n=1828)

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)***:**
Using the 2009 data, we will examine the correlation of each of the composites with the global ratings as a measure of criterion validity. We will also examine scaling success for each composite.

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)***:**
Analysis of 2009 data is currently being conducted.

For 2005 data, Table 27c (mail sample) on page 85 of Harvard Final report shows correlation of the 5 Composites with the global items in the survey)

**2c**
C☐
P☐
M☐
N☐

**2d. Exclusions Justified**

2d.1 **Summary of Evidence supporting exclusion(s)**:
AHRQ will harmonize this measure with other short stay measures.
Those who were discharged to another care facility and not discharged home were excluded because they may not keep their experience in the two different settings distinct. Deceased persons can not respond to

**2d**
C☐
P☐
M☐
N☐
NA☐

**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.

**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 ... [4]

**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.

survey and proxy should not be allowed to be included with other discharged residents.

2d.2 **Citations for Evidence:**
expert opinion and common sense (e.g., deceased)

2d.3 **Data/sample** *(description of data/sample and size)*: 2005 field test data and 2009 field test data

2d.4 **Analytic Method** *(type analysis & rationale)*:
frequency analysis of field test data

2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:
based on 2005 data, almost half (45%) were found ineligible based on being discharged to another facility; about 27% were found to be deceased; about 21% were found ineligible based on less than 5 or more than 90 days of a stay.

Analysis of 2009 data currently being conducted.

**2e. Risk Adjustment for Outcomes/ Resource Use Measures**

2e.1 **Data/sample** *(description of data/sample and size)*: 2009 field test data from 60 nursing homes in Maryland (n=1828); the 2005 data is not large enough for risk analysis.

2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*:
We will use the risk adjustment analysis method used for the Family Member Survey on pages 26-33 of AIR Final Report

2e.3 **Testing Results** *(risk model performance metrics)*:
Analyses of 2009 data are currently being conducted

2e.4 **If outcome or resource use measure is not risk adjusted, provide rationale:**

**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)*: 2005 field test data from 12 nursing homes in four New England states (n=127) and 2009 field test data form 60 nursing homes in Maryland (n=1828)

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
For statistical significance for each composite or global item, we will use a t-test comparing each nursing home mean to the mean of all the nursing home means.

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)*:
from the 2005 data:
The mean and standard deviations (SD) for the composites for the discharged resident survey (mail sample) are:
Composite 1: Environment -- mean = 5.47 (1.14)
Composite 2: Care-- mean = 6.97 (1.79)
Composite 3: Communication and Respect- mean= 8.47 (1.67)
Composite 4: Autonomy – mean= 2.82 (0.34)
Composite 5: Activities – mean = 2.50 (0.67)

Additional statistical detail on pages 84 of Harvard Final Report

The 2009 data (N=1828) is currently being analyzed and these 5 composites may change based on these analyses.

**2f**
**C**☐
**P**☐
**M**☐
**N**☐

**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.

| | |
|---|---|
| **2g. Comparability of Multiple Data Sources/Methods**<br>2g.1 **Data/sample** *(description of data/sample and size)***:** not available<br>2g.2 **Analytic Method** *(type of analysis & rationale)***:**<br>2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)***:** | **2g**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐<br>**NA**☐ |
| **2h.** D**isparities in Care**<br>2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)***:** not applicable<br>2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:** | **2h**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐<br>**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?**<br>**Rationale:** | *2*<br>**C**☐<br>**P**☐<br>**M**☐<br>**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**<br>3a.1 **Current Use:** Testing not yet completed<br>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). <u>If not publicly reported</u>, state the plans to achieve public reporting within 3 years*)**:**<br>Maryland Health care Commission (MHCC) plans to publicly report the results of the Short Stay Resident Instrument within 3 years. This public reporting effort would be similar to what MHCC does now for their family member nursing home survey.<br>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). <u>If not used for QI</u>, state the plans to achieve use for QI within 3 years***):**<br>MHCC plans to provide feedback to the nursing homes who participated in the 2009 field test concerning their results from the Short Stay Resident Instrument once AHRQ has completed analyses. This quality improvement effort would be similar to what MHCC does now for their family member nursing home survey.<br>**Testing of Interpretability** (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)<br>3a.4 **Data/sample** *(description of data/sample and size)***:** six focus groups in 3 states - four with nursing home residents and two with family members<br>Cognitive testing:<br>Round 1: 52 residents in 5 homes<br>Round 2: 15 residents in 3 homes<br>Round 3: 19 residents in 3 homes<br>Round 4: 27 residents in 3 homes<br>Round 5: 31 residents in 3 homes | **3a**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐ |

**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.

**Comment [KP22]: 3a**. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for <u>both</u> public reporting (e.g., focus group, cognitive testing) <u>and</u> 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.

Round 6: 16 residents in 2 homes
Round 7: 19 residents in 2 homes

**3a.5 Methods** *(e.g., focus group, survey, QI project)***:**
Six focus groups were conducted with residents and family members and there were 7 Rounds of cognitive testing between 2001 and 2005. We conducted a pretest in one nursing home in May 2005.

Because short stay residents' stays are usually primarily a medical event, not a total living experience, we used expert opinion to adapt the survey instrument for long stay residents slightly (adding a question about therapy received and deleting questions about hearing aids, and eye and dental care). We also adapted the questions for discharged patients ("when you were in the nursing home...") and for a mail survey form since analysis of national MDS data showed that cognitive impairment was less prevalent among short-stay than long stay residents, making a mail survey feasible.

**3a.6 Results** *(qualitative and/or quantitative results and conclusions)***:**
Focus groups results: Resident focus groups indicated that issues of greatest concern were cleanliness of the facility, noise, food, training,competency of staff, language issues, continuity of staff and receiving correct medication. Some issues suggested in the literature, such as "safety" were not considered as important tour participants. Likewise, many participants reported the CAHPS domain "communication with doctors" as being irrelevant to their quality of care (QoC) because they did not see doctors as often as other staff. Since CAHPS was originally created for use in ambulatory settings, it makes sense that some domains are inappropriate for nursing home residents. Much of what was learned in the resident focus groups was echoed in the family groups. The main concerns of the family groups were cleanliness, availability of activities, and adequacy and respectfulness of staff. Concerns about medical care were much less important to both groups than day-to-day activities. We also learned from the family groups that they may not be as knowledgeable proxy responders for the care of nursing home residents.

Cognitive Testing of Resident Instrument
Using the information from the focus groups and literature review, we drafted an instrument. We then conducted a series of cognitive interviews to ensure that candidate survey items were understood in a consistent way by respondents as well as to learn whether the respondent’s as well as to learn whether the respondent’ answers accurately reflected what they have to say on the topic. Interviewers followed a semi-structured protocol, which included the survey questions and a set of scripted cognitive probes about each question. The protocol called for interviewers to ask the test questions as worded, obtain answers to one or a short series of questions, then proceed to the cognitive probes. The team used professional interviewers to conduct a total of seven rounds of cognitive interviews. Again, nursing homes near the research organizations were recruited by letters and personal contact with researchers.

Round I. Nursing homes provided a list of both long-and short-stay residents. They were asked to include residents who they felt could answer our interviewer-administered questionnaire, some who would probably have difficulty but could do so (those with some difficulty in daily decision making or who cannot always make themselves understood), and some who were unlikely to be able to complete the process (those with memory problems, more sever problems with daily decision making, or who often have difficulty making themselves understood). Interviewing teams talked to the residents on the list, explained the study and the interview process, and then administered a short cognitive screener. The screener consisted of eight items drawn from a variety of other screeners intended to test orientation, recall, and reasoning. If a respondent answered six or more questions correctly, he or she was eligible to be interviewed in this round. Very few of the respondents failed this cognitive screener.

The goal of the first round was to evaluate the specific wording and concepts in the draft survey. A particular focus of the testing was whether respondents could handle a four-category response task (always, usually, sometimes, never) or if a two-category response task (yes or no) worked better. After testing, we found that the “always” to “never” response task, one of the cores response tasks for CAHPS instruments, was very difficult for nursing home residents. The dichotomous response choice (“yes” or “no”) as an alternative did not work well either. Leaving out any sense of frequency in a question such as “(In the last week), did you get help washing your face or combing your hair?” makes the question ambiguous rather than making it simple. Respondents were unsure whether the question asking whether one always got help or ever got help. Some sense of frequency was essential to making the question comprehensible and the answers meaningful.

We also found that respondents, in general, paid almost no attention to the time frames in the questions. Yet, perhaps the most important thing the team learned from the cognitive interviews was that summarizing across time and people was a major challenge for respondents. When we e asked respondents how they decide on their answers, we found that there was a tendency for them to simplify the cognitive task by focusing on a single individual or a single event, thereby making the tasks easier. In many of the Quality of Care (QOC) items, the events asked about occur frequently and thus do not stand out as events very much. For example, thinking about all the times in the last week that eating or going to the bathroom occurred was very hard for respondents to synthesize. They were clearly unable to figure out how often these very common occurrences happened, let alone how many of those times they had problems.

Rounds 2 and 3. After the first round of cognitive interviewing, the team realized that before concentrating on question content, they first had to figure out what type and form of question most nursing home residents could answer. We determined that there were three key features that could vary in questions to measure nursing home experiences:

- Type of question, for example, report (occurrence or frequency of event), or rating (resident's perception of event)
- Time period asked about, for example, single day, multiple days, non-specific time period
- Type of response task:

Reports, for example, Yes/No; frequency reports (e.g., " always" to "never"); or days-based frequency (e.g., "every day, some days, no days")
Ratings, for example, ordered adjectives (e.g., "every day, some days, no days")
Ratings, for example, ordered adjectives (e.g., "excellent" to "poor"); Comparative evaluation, numbered rating.

We decided to take a few concepts (such as food, getting help, and noise) and develop alternatives that varied all the question characteristics listed. By creating a taxonomy of possible options, the team was able to test many different ways to ask these questions. Appendix B in the Journal of Aging and Social Policy article shows an example of the different questions that could be asked about one concept. These variations were then used in the next rounds of cognitive interviewing. The goal of both rounds two and three was a systematic test of how best to get information from nursing home residents.
In these two rounds, the sample again was based on suggestions from the nursing home staff and a score of 6 to 8 on the cognitive screener. With respect to time period, the team found that asking about "yesterday" did not work well because it provided a very limited basis for respondents to report. Also, some respondents answered about the last time an event did occur (even if it did not happen on the day in question). The phrase "last week" was problematic, since respondents had difficulty summarizing over time and focusing on a specific reference period. The non-specific present (asking about "how things are going now") provided the most reliable responses, based on respondents' descriptions of how they decided on their answers.
In terms of the type of questions, the team found that asking for a rating was easier than asking for a report of the same thing (since ratings do not rely on a respondents having to summarize their experiences). For example, asking residents to report on how often they liked the food at the nursing home was much more difficult than asking them to rate the food. Ratings tended to reflect residents' overall descriptions of care in particular areas without requiring them to integrate multiple discrete events.
Knowing which question type seemed to work better, then team refined the testing to focus on the various types of response tasks. The team tested ratings with adjectives, numbered rating scales, and comparative evaluations and found that adjectives, numbered rating scales, and comparative evaluations and found that adjective scales (e.g., "excellent" to "poor") were harder for respondents to remember and use, even when the responses were listed on a show cared. They also had trouble with the comparative evaluations. Whether the scales were difficult, or whether the comparative evaluation concept was cognitively complex for people is uncertain. Testing showed that they best form for most QOC questions was to ask ratings in the non-specific present using a 0 to 10 rating scale. Residents reported more comfort and ease with using numbers 0 to 10 then using the given worded response categories. Using numbers simplified the response task, and residents were not distracted by the meaning or emotional content of the words. Respondents could explain their answer choices and what higher or lower scores would signify.
Round 4 or 5. Once the question format was decided, these rounds focused on question content and

wording. In contract to the previous rounds, no cognitive screener was used to eliminate residents from the sample; residents were chosen from a census list of current residents (with their CPS scores) provided by the nursing home. Researchers attempted to interview a mix of those with high and low CPS scores (ranging form 0 to 5). These rounds also tested the vignettes (described later in this article). Testing rounds 6 and 7 are described as follows.

Merging QOC and Quality of Life (QOL) Constructs
When CMS decided it wanted the nursing home resident experience instrument to have both QOC and QOL combined, the team had to select and, if needed, modify QOL items developed by Kane and colleagues and to merge them with the NHCAHPS QOC items. First, the team compared the domain and item content of the QOL measures to that of a variety of CAHPS measures that were currently under development for patient populations that are frail and require intense care included questions used in the hospital CAHPS instrument, the in-center hemodialysis CAHPS instrument, and, of course, the fall 2003 NHCAHPS instrument. This analysis revealed that may of the items included in QOL measures actually tapped QOC and were very similar to the items included in other CAHPS instruments. That is, even though the domains in QOL instruments referred to aspects of QOL (e.g., autonomy, dignity), the actual items included in some of those domains referred to QOC (e.g., whether care providers communicated with courtesy and respect). The concept of “QOL” is very broad and is approached from various disciplines and perspectives. But what all approaches have in common is the idea that QOL is a subjective state of being. On the other hand, QOC is a report of one’s experience of the care delivered. The two concepts are often distinguished by saying that QOC refers to health care process (activities of delivering care) and QOL refers to health care outcomes (the subjective state of the person to whom care was delivered as an outcome of care processes).
The team systematically reviewed the content of the QOL items to determine whether it was unique to QOL (e.g., autonomy, spiritually) or whether it referred to QOC (e.g., communication with staff). The ultimate goal was to identify content that should be used to supplement the NHCAHPS OQC items and to identify items that referred uniquely to QOL for inclusion in the NHCAHPS survey. To help decide which QOL items to include, the team sued several criteria, including whether the item was actionable for nursing home quality improvement, what the response distribution looked like, what the item’s relationship was to other variables and to overall QOL rating, and whether the item was able to discriminate among nursing homes.
In rounds 6 and 7 of cognitive interviewing, we focused on the QOL questions to determine residents’ understanding of the new items and various response tasks. In addition, we tested if there are any order effects of QOL items and selected QOL items. Some of key findings learned from cognitive testing the QOL items were:

• Response tasks. The 0 to 10 rating scale (worse possible to best possible) did not work for many QOL items. “Mostly yes/Mostly no” also was not an adequate response task for respondents. “Yes/No/Sometimes” was tested and found to be preferable for QOL items.
• Order/Structure. Respondents found it cognitively complex to switch back and forth between the 0 to 10 scales and the “Mostly yes/Mostly no” questions, even if question content was similar. The order of the questions was changed to pull all the 0 to 10 rating questions first. This worked much better for respondents in round 7 of testing.
• Screeners. Some items that we thought all respondents could answer, such as being left lying in one position so long that it hurt, actually needed screening questions(e.g., first asking if no could turn/move oneself in bed).
Additional edits were made, based on the cognitive testing results of the QOL items. In May 2005, a pretest of the merged questionnaire was completed. The pretest provided information about how the final combined instrument worked together as well as providing some information about the actual protocol used in the field test.

Summary of Lessons learned from cognitive testing: The resident NHCAHPS developed demonstrates the critical role of cognitive interviewing to test survey items with the intended respondents prior to full-scale implementation, particularly for a population with cognitive challenges, such as nursing home residents. The cognitive testing results helped the team understand the most appropriate wording for items, as well as provide guidance on types of questions, time period asked about, and type of response task. In contrast to other CAHPS surveys, the NHCAHPS team concluded that ratings were more useful than reports because of the difficulty that residents had with summarizing over time and people. Because of repeated evidence that residents had trouble with reference periods, our recommendations is to use the non- specific present,

| | |
|---|---|
| in contrast with typical survey methodology and other CAHPS surveys where explicit time reference periods are used. The NHCAHPS testing found that 0 to 10 response scale appeared to work well with nursing home residents for many of the QOC questions. This use of 0 to 10 scales is consistent with other CAHPS surveys and some other research with elderly. Our testing did find, however, that a different response scale (yes/sometimes/no) was needed for many of the QOL items. | |
| **3b/3c. Relation to other NQF-endorsed measures**<br>3b.1 **NQF # and Title of similar or related measures:** | |
| (for NQF staff use) **Notes on similar/related** endorsed **or submitted measures:** | |
| **3b. Harmonization**<br>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):<br>3b.2 **Are the measure specifications harmonized? If not, why?** | **3b**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐<br>**NA**☐ |
| **3c. Distinctive or Additive Value**<br>3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**<br>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:**<br>This will be the only patient experience measure for short stay nursing home residents. | **3c**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐<br>**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?**<br>**Rationale:** | **3**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐ |
| **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**<br>4a.1-2 **How are the data elements that are needed to compute measure scores generated?**<br>Survey | **4a**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐ |
| **4b. Electronic Sources**<br>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*)<br>No<br>4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**<br>this is a survey measure so electronic capture is not part of design | **4b**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐ |
| **4c. Exclusions**<br>4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** | **4c**<br>**C**☐<br>**P**☐<br>**M**☐ |

**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.

**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.

No

4c.2 **If yes, provide justification.**

N☐
NA☐

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

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

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.**

There could be issues if the entity collecting the data does not follow the guidelines for survey administration (e.g.,drawing the sample and assuring confidentiality). Unless the sponsor permits direct access to the resident records for random sampling, it is possible that the nursing home may select discharged residents likely to give more favorable responses (or exclude those likely to give unfavorable responses) when selecting records for the sample. In addition, errors could be introduced if an entity adds non-Nursing Home CAHPS items before any of the core survey questions in the Nursing Home CAHPS Discharged Resident Survey. The core survey items are all those questions prior to the "About You" section of the survey. AHRQ has a CAHPS User Group support contract that is available to provide technical assistance for entities wishing to implement this survey- this can help reduce errors.

**4d**
C☐
P☐
M☐
N☐

**4e. Data Collection Strategy/Implementation**

**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).

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:**

Lessons learned from 2005 CAHPS Field test:

DESIGN & PROCEDURES

Short-term nursing home residents' stays are usually primarily a medical event, not a total living experience. The cognitive problems that are so prevalent among long-term residents are much less prevalent in the short-term population. The survey instrument for long-term residents was adapted slightly (adding a question about therapy received and deleting questions about help with hearing aids, and eye and dental care) and put into a mail survey form for the short-term residents. For convenience and timing, the short-term residents of the nursing homes participating in the long-term field test were used as the sample frame for this effort.

At the same time the nursing homes participating in the long-term resident survey were providing data about their long-terms residents, they were also asked to provide a list of all those who had been discharged from the nursing home within the past 2 months.

SAMPLE:

Eligible respondents were residents who had been in the home for at least 5 days and who had not been in the home for more than 90 days and who were not deceased or discharged to another care facility. We excluded residents whose most recent MDS assessment indicated that they were "severely impaired in cognitive skills for daily decision making", that they were comatose, and those for whom a legal guardian was required to make medical decisions were excluded.

PROTOCOL:

The protocol was to send an initial mailing, with a cover letter, fact sheet, copy of a self-administered questionnaire, and a postage-paid return envelope. A packet with a second questionnaire was to be sent after two weeks to the non-responders. Then, nonresponders were to get a reminder call to make sure they had received the questionnaire, answer any questions, and urge participation. The first mailing was on June 28. Unfortunately, 5 nursing homes could not provide their completed samples until well into July. Because of a hard deadline for data collection of mid-August, there was not time for follow-up mailings to those nursing homes. Those homes received their first (and only) mailing at the same time as the other 6 received their second mailing. One nursing home was not recruited until the end of July. Although in-person interviews were collected, this home did not participate in the mail portion of the field test. CSR called everyone who had not yet returned a completed questionnaire to prompt people to return the written survey. Telephone reminder calls were conducted by professional interviewers. Each case had up to 3 calls made on different days and at different times of the day to attempt to make contact with the appropriate respondent.

**4e**
C☐
P☐
M☐
N☐

FINDINGS:
1. Sampling and Eligibility
The initial sample consisted of 381 residents from 11 nursing homes. 133 residents were ineligible because they did not meet the eligibility criteria. The major reason for ineligibility was being discharged to another care facility or being deceased. Very few in the sample were ineligible because of cognitive impairment or having a legal guardian/legal oversight. In order to obtain sufficient levels of response for reporting results by individual nursing home, future research might consider different sampling options, including using a rolling sample, where residents over several months are surveyed, rather than just the 2 month window that was used for the field test.
2. Data Collection
We found that the quality of the contact information was generally fairly good. About 70% of the sample sent from the nursing homes included phone numbers. Only one nursing home was unable to provide any phone numbers. Overall, almost 52% of the eligible sample returned a completed survey. With time to implement a good mail protocol (including an option of telephone interview), this field test experience would suggest acceptable rates of return could be achieved. Discharged residents who received the more standard 2-mailings had a 57% response rate, while those that only received one mailing had a 43% response rate. There were very few explicit refusals to participate.

4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**
This CAHPS survey instrument and all composite measures are in the public domain and free to use. The costs associated with implementing these measures are the cost of data collection, analysis and facility feedback or public reporting.

The 2009 Maryland Health Care Commission(MHCC)paid approximately $25.37 per completed mail survey of short stay residents (n=1828 completes)for the costs of data collection and data entry(analysis costs are paid by AHRQ). The mail protocol was 2 rounds of mail with a reminder phone call but no phone interview followup.

The total direct costs of the 2005 CAHPS pliot test, not including overhead or sampling costs, for the mail portion of the study cost approximately $2000 – about $182 per home (for 11 homes) or about $16.25 per completed survey. The costs of doing this mail study are not a good basis for estimating future costs. Because of time constraints, a complete mail with phone call reminder protocol was not implemented. Also, the sampling for the in-person and mail surveys was done together, so the mail survey costs for sampling cannot be broken out separately.

4e.3 **Evidence for costs:**
Fall 2009 field test with MHCC and May 2005 field test in 4 states in New England

4e.4 **Business case documentation:** The intent of the NHCAHPS initiative (also known as Nursing Home CAHPS) is to provide a set of standardized survey instruments and data collection methodology for measuring residents' (both long - and short-stay) and families' perspectives on nursing home care. While many nursing homes may currently collect information on patient satisfaction, prior to NHCAHPS there has been no national standard for collecting or publicly reporting nursing home residents' and families' perspectives of care information that would enable valid comparisons to be made across all nursing homes.

In order to make "apples to apples" comparisons to support consumer choice, AHRQ has recognized the importance of creating a standard measurement approach. NHCAHPS is a core set of questions that can be combined with a broader, customized set of nursing home-specific items. NHCAHPS survey items complement the data a nursing home may currently collect to support improvements in internal customer services and quality related activities. For many nursing homes, short stay residents represent a significant part of their population.

Three broad goals have shaped the NHCAHPS survey. First, the survey is designed to produce comparable data on the nursing home residents' and family members' perspective on care that allows objective and meaningful comparisons between nursing homes on domains that are important to them. Second, public reporting of the survey results is designed to create incentives for nursing home to improve their quality of care. Third, public reporting will serve to enhance public accountability in health care by increasing the

| | |
|---|---|
| transparency of the quality of nursing home care provided in return for the public investment. Because the government (federal and state combined) pays for almost two-thirds of the $131 billion of total nursing home costs (2008 statistics), the Centers for Medicare & Medicaid Services (CMS) are interested in the consumers' perspective on the quality of care they receive. As the federal agency responsible for nursing home quality oversight, CMS has supported the development of a consumer experience survey for both residents and their family members. With these goals in mind, the NHCAHPS project has taken substantial steps to assure that the survey is credible, useful, and practical. This methodology and the information it generates is available to the public. | |
| **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?**<br>**Rationale:** | **4**<br>**C**☐<br>**P**☐<br>**M**☐<br>**N**☐ |
| **RECOMMENDATION** | |
| (for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.** | **Time-limited**<br>☐ |
| **Steering Committee: Do you recommend for endorsement?**<br>**Comments:** | **Y**☐<br>**N**☐<br>**A**☐ |

## CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
Agency for Healthcare Research and Quality , 540 Gaither Road , Rockville, Maryland, 20850

Co.2 **Point of Contact**
Judith, Sangl, Sc.D., jsangl@ahrq.gov , 301-427-1308-

**Measure Developer If different from Measure Steward**
Co.3 **Organization**
Agency for Healthcare Research and Quality , 540 Gaither Road , Rockville, Maryland, 20850

Co.4 **Point of Contact**
Judith, Sangl, Sc.D., jsangl@ahrq.gov , 301-427-1308-

Co.5 **Submitter If different from Measure Steward POC**
Judith, Sangl, Sc.D., jsangl@ahrq.gov , 301-427-1308-, Agency for Healthcare Research and Quality

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.**
The development of the NHCAHPS resident instrument was a multi-phase process. In the initial phase, CMS requested AHRQ and the CAHPS team to investigate the methodological challenges of conducting a survey with nursing home residents. This phase examined sampling issues, cognitive screeners, data collection methods, and possible survey content. The CAHPS team conducted interviews on these topics with the following experts: Steve Albert, Kitty Buckwalter, Tim Case, Ann Gruber-Baldini, Catherine Hawes, Ted Johnson, Rosalie Kane, Powell Lawton, Vince Mor, John Morris, Peter Norton, Sandra Simmons, Phil Sloan, Joan Teno, Gwen Uman, Sheryl Zimmerman, and Jackie Zinn. AHRQ and the CAHPS team convened a Methodological Expert Group (MEG) to further explore these issues. The MEG included: Robert and Rosalie Kane; Farida Ejaz, Catherine Hawes; Kathleen Buckwalter; Andrew Kramer; Powell Lawton; Jay Magaziner; Vincent Mor; Rudolph Moos; John Schnelle; Philip

Sloane; Liane Soberman; Joan Teno; and Sheryl Zimmerman. At the end this initial Phase, CMS, AHRQ, and the CAHPS team concluded that it was feasible to obtain reliable reports of experiences in the nursing home from many long stay nursing home residents by conducting in-person surveys. AHRQ also had extensive consultations with CMS and the Kanes when working on the merger of the Quality of Life items with the Quality of Care items.

Ad.2 **If adapted, provide name of original measure:**
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:** 2007
Ad.7 **Month and Year of most recent revision:**
Ad.8 **What is your frequency for review/update of this measure?** 2nd field test done fall 2009 to get larger sample size for analyses
Ad.9 **When is the next scheduled review/update for this measure?** 01, 2011

Ad.10 **Copyright statement/disclaimers:** CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. This CAHPS® questionnaire should be used without modification to the core set of questions. Supplemental questions may be added after the core set of questions and before the demographic question section. Please consult Guidelines for Modifying and Naming CAHPS Surveys at https://www.cahps.ahrq.gov/content/products/PROD_ModifySurveys.asp

Ad.11 -13 **Additional Information web page URL or attachment:** Attachment Integrated NH CAHPS Report-1-19-06-634092527834729936.pdf

**Date of Submission** (*MM/DD/YY*)**:** 10/07/2010

**Page 4: [1] Comment [k4] Karen Pace 10/5/2009 8:59:00 AM**

**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.

**Page 4: [2] Comment [k5] Karen Pace 10/5/2009 8:59:00 AM**

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.

**Page 4: [3] Comment [k7] Karen Pace 10/5/2009 8:59:00 AM**

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.

**Page 7: [4] Comment [KP14] Karen Pace 10/5/2009 8:59:00 AM**

**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,