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 <mark>yellow highlighted</mark> 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 (<mark>yellow</mark> highlighted areas).*

Steering Committee: Complete all <mark>pink</mark> 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)

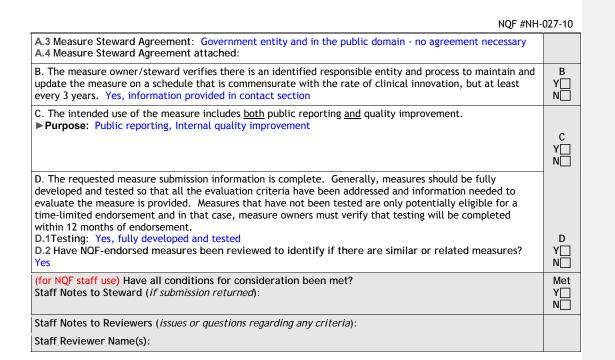| (for NQF staff use) NQF Review #: NH-027-10   NQF Project: Nursing Homes 2010 |
|---|
| **MEASURE DESCRIPTIVE INFORMATION** |
| **De.1** Measure Title: Consumer Assessment of Health Providers and Systems (CAHPS®) Nursing Home Survey: Long-Stay Resident Instrument |
| **De.2** Brief description of measure*: The CAHPS® Nursing Home Survey: Long-Stay Resident Instrument is an in-person survey instrument to gather information on the experience of long stay (greater than 100 days) residents currently in nursing homes. The Centers for Medicare & Medicaid Services requested development of this survey, and can be used in conjunction with the CAHPS Nursing Home Survey: Family Member Instrument and Discharged Resident Instrument. The survey instrument provides nursing home level scores on 5 topics valued by residents: (1) Environment; (2) Care; (3) Communication & Respect; (4) Autonomy and (5) Activities. In addition, the survey provides nursing home level scores on 3 global items. |
| **1.1-2** Type of Measure: Patient experience
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
De.5 IOM Quality Domain: Patient-centered
De.6 Consumer Care Need: |

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* **A.1** Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)? Yes **A.2 Indicate if** Proprietary Measure (*as defined in measure steward agreement*):	A Y ☐ N ☐

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

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) 1a. High Impact	**Eval Rating**
(for NQF staff use) Specific NPP goal:	
1a.1 Demonstrated High Impact Aspect of Healthcare: Patient/societal consequences of poor quality **1a.2** **1a.3** Summary of Evidence of High Impact: According to the 2004 National Nursing Home Survey (NNHS), there were approximately 1.5 million nursing home residents in 16,100 nursing home facilities (Jones et al, 2009). They are a population with significant limitations in activities of daily living (ADLs) with 51% receiving assistance with all 5 ADLs (bathing, dressing, toileting, transferring or eating)and less than 3% receiving no ADL help (Jones et al 2009); about 69% have cognitive impairment as measured by the Cognitive Performance Scale (CMS 2008). The National Health Expenditures Accounts(CMS, 2009) estimate that nursing home costs totaled $131 billion in 2008. 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. Concurrent with changes from OBRA'87 implementation, a radical rethinking of the long term care system	**1a** C☐ P☐ M☐ N☐

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

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

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?id=596) also includes patient and family engagement as one of its priorities.

1a.4 Citations for Evidence of High Impact: Jones, A. L., Dwyer, L.L. , Bercovitz, A.R., Strahan, G. The National Nursing Home Survey: 2004 Overview. National Center for Health Statistics. Vital Health Stat. 13(167). 2009

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 vulnerable 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.4 Summary of Data on disparities by population group:

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.

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: Observational study, Expert opinion

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 statisfaction 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:
not applicable

1c
C ☐
P ☐
M ☐
N ☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   4

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 Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  o 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 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]]

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	1
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, met? Rationale:	1 Y☐ N☐

2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating

2a. MEASURE SPECIFICATIONS	

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

2a. Precisely Specified

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome*):
The following topics are measured for nursing homes from a resident's perspective:
Composite 1: Environment – sum of applicable resident scores on 8 survey items (see codebook for points assigned to each response category) related to aspects of environment in nursing home
Composite 2: Care - sum of applicable resident scores on 5 survey items
Composite 3: Communication and Respect- facility score is sum of applicable resident scores on 3 survey items
Composite 4: Autonomy - sum of applicable resident scores on 3 survey items
Composite 5: Activities – sum of applicable resident scores on 2 survey items
Global Items:
Global Rating of care received from staff: sum of resident scores on 0 to 10 scale
Global Rating of overall nursing home: sum of resident scores on 0 to 10 scale
Global item whether respondent would recommend nursing home: sum of resident scores on item (see codebook for points assigned to each response category)

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):
non-specific present – see 3a.6 for cognitive testing results for this time window decision

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):
(Note: Question # is from final survey which may differ from pilot survey)
Composite 1: 8 survey items Q1, Q3, Q4, Q5, Q6, Q18, Q19, Q20
Composite 2: 5 survey items Q8, Q9, Q10, Q12, Q29
Composite 3: 3 survey items Q13, Q14, Q15
Composite 4: 3 survey items Q30, Q31, Q32
Composite 5: 2 survey items Q33, Q34
Global items: 3 survey items Q16, Q17, Q35

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*):
The denominator is the total number of surveys for respondents that meet CAHPS completion standard and any applicable screener (discussed in details below)

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 18+

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the*

2a-specs
C☐
P☐
M☐
N☐

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

denominator):
non-specific present – see 3a.6 for cognitive testing results for this time window decision

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*):
Composite 1: Environment
the denominator is the total number of completed surveys for 7 out of 8 questions in this composite excluding Q3, where it is the number of surveys completed by all those who responded "yes" to screener Q2
Composite 2: Care
the denominator is the total number of completed surveys for 2 out of 5 questions in this composite excluding these questions:
Q8: the number of surveys completed by all those who responded "yes" to screener Q7
Q12: the number of surveys completed by all those who responded "yes" to screener Q11
Q29: the number of surveys completed by all those who responded "yes" to screener Q28
Composite 3: Communication and Respect
the denominator is the total number of completed surveys for all 3 questions
Composite 4: Autonomy: the denominator is the total number of completed surveys for all 3 questions in this composite
Composite 4: Activities: the denominator is the total number of completed surveys for the 2 questions in this composite
Global Items: for all 3 global items the denominator is the total number of completed surveys.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*): We exclude residents who (1)are under age 18, (2)are comatose, (3) are severely impaired in mental status or in cognitive skills for daily decisionmaking, (4) cannot answer 3 questions in a row; (5) conscious but unresponsive to interviewer and (6) unable to speak English for survey (we expect Spanish translation will be available in summer/fall 2011). All residents whose length of stay (LOS) in the facility is equal to or less than 100 days from the date of admission will also be excluded. Residents who return to the nursing home following any hospital discharge will not have their stay reset to zero when they return to the facility. AHRQ will harmonize its specification on long stay residents with CMS.

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*):
1. Residents who are under age 18.
2. Residents whose last MDS 3.0 evaluation had a Brief Interview for Mental Status(BIMS)score of less than 8 on item C0500 or the staff assessment of mental status indicated they were "severely impaired in cognitive skills for daily decision making" (MDS 3.0 C1000 = 3).
3. Residents who were in a coma (MDS3.0 B0100 =1).
4. Residents who had not been in the home for more than 100 days or would not be by the time of data collection/date of interview. Residents who return to the nursing home following any hospital discharge will not have their stay reset to zero when they return to the facility. AHRQ will harmonize its specification on long stay residents with CMS

During survey administration there were the following additional exclusions determined by trained interviewers:
1. Non-English speaking (pilot survey only available in English)- (we expect Spanish translation will be available in summer/fall 2011).
2. unable to answer 3 questions in a row
3. unresponsive to interviewer

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: 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*):

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.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*):
SCORING FOR CAHPS NURSING HOME SURVEY: LONG STAY RESIDENT INSTRUMENT

1. Global ratings and items
•   Measured by resident's overall care from staff on a scale of 0-10 (Q16)
•   Measured by resident's overall rating of the nursing home on a scale of 0-10 (Q17)
•   Measured by whether the resident would recommend the nursing home to others on a scale of
Definitely No, Probably No, Probably Yes, and Definitely Yes (Q35)

2. Domains of care
•   Environment (Q1, Q3, Q4, Q5, Q6, Q18, Q19, Q20)
•   Care (Q8, Q9, Q10, Q12, and Q29)
•   Communication/Respect (Q13, Q14, & Q15)
•   Autonomy (Q30, Q31, & Q32)
•   Activity (Q33 &Q34)

3. Production of Nursing Home scores – Global items

•   Nursing home level ratings for Q16 and Q17 are presented using a three-category display for the 0-
10 scale question: 0-6, 7-8, and 9-10.
•   Q35: Nursing home level scores are presented using percentages for the following three categories:
definitely would recommend, probably would recommend, and probably not or definitely not recommend.

4. Production of Nursing Home scores - Domain-level composites
There are five domain-level composites included in the Nursing Home Long-Stay Resident Questionnaire:
Environment, Care, Communication/Respect, Autonomy, and Activities.

•   Environment
     ursing home score for this composite is produced by combining responses to eight questions:
o   Q1: "What number would you use to rate the food here at the nursing home?"
o   Q3: "When you eat in the dining room, what number would you use to rate how much you enjoy
mealtimes?"
o   Q4: "What number would you use to rate how comfortable the temperature is in the nursing
home?"
o   Q5: "What number would you use to rate how clean the nursing home is?"
o   Q6: "What number would you use to describe how safe and secure you feel in the nursing home?"
o   Q18: "Is the area around your room quiet at night?"
o   Q19: "Are you bothered by noise in the nursing home during the day?" (note: "No" represents
higher quality so this question needs to be reverse coded)
o   Q20: "If you have a visitor, can you find a place to visit in private?"

•   Respondents to five of the above questions can answer on a 0-10 scale. Respondents to three of
the above questions can answer "yes", "no" or "sometimes" to each. A nursing home's score on the
"Environment" composite is the proportion of cases in each response category.

The steps to calculate a nursing home provider's composite score follow:

Step 1 – Calculate the proportion of cases in each response category for the first question:
P11 = Proportion of respondents who gave a rating of "0 to 6"
P12 = Proportion of respondents who gave a rating of "7 or 8"
P13 = Proportion of respondents who gave a rating of "9 or 10"

Follow the same steps for the second question:
P21 = Proportion of respondents who gave a rating of "0 to 6

P22 = Proportion of respondents who gave a rating of "7 or 8"
P23 = Proportion of respondents who gave a rating of "9 or 10"

Repeat the same procedure for each of the rating questions in the composite.

For the three questions with "yes/no/sometimes; consider "yes' to be equivalent to rating of "9 or 10"; "sometimes to be equivalent to rating of "7 or 8" and " no" to be equivalent to rating of "0 to 6", except for Q19 where it would be reverse coded because "no" represents better quality.

Survey sponsors may choose alternative methods to combine proportions (such as different groups of rating from 0 to 10).

Step 2 – Combine responses from the questions to form the composite
Calculate the average proportion responding to each category across the questions in the composite. For example, in the "Environment" composite (eight questions), calculations would be as follows:

PC1 = Composite proportion who said "yes" or gave a rating of "9 or 10" = (P11 + P21 + P31 + P41 + P51 + P61 + P73* + P81) / 8
* Q19 is reverse coded

PC2 = Composite proportion who said "sometimes" or gave a rating of "7 or 8"= = (P12 + P22 + P32 + P42 + P52 + P62 + P72 + P82) / 8

PC3 = Composite proportion who said "no" or gave a rating of "0 to 6"= (P13 + P23 + P33 + P43 + P53 + P63 + P71* + P83) / 8
* Q19 is reverse coded

• Care
The nursing home score for this composite is produced by combining responses to five questions:
o Q8: "What number would you use to rate how well the medicine worked to help with aches or pain?"
o Q9: "What number would you use to rate how well the staff help you when you have pain?"
o Q10: "What number would you use to rate how quickly the staff come when you call for help?"
o Q12: "What number would you use to rate how gentle the staff are when they're helping you?"
o Q29: "Do the staff make sure you have enough personal privacy when you dress, take a shower, or bathe?"

Respondents to four of the above five questions can answer on a 0-10 scale. Respondents can answer "yes," "no," "sometimes," to one question - Q29. The steps to calculate a nursing home's composite score for this domain are similar to Environment composite except that in Step 2, each composite proportion category would be divided by 5 (the total number of items).

• Communication/Respect
The nursing home score for this composite is produced by combining responses to three questions:
o Q13: "What number would you use to rate how respectful the staff are to you?"
o Q14: "What number would you use to rate how well the staff listen to you?"
o Q15: "What number would you use to rate how clearly the staff explain things about your care to you?"

Respondents to the above questions can answer 0-10 to each. The steps to calculate a nursing home's composite score for this domain are similar to Environment composite except that in Step 2, each composite proportion category would be divided by 3 (the total number of items).

• Autonomy
The nursing home score for this composite is produced by combining responses to three questions:
o Q30: "Can you choose what time you go to bed?"
o Q31: "Can you choose what clothes you wear?"
o Q32: "Can you choose what activities you do here?"

Respondents to the above questions can answer "yes", "no" or "sometimes" to each. The steps to calculate a nursing home's composite score for this domain are similar to Environment composite except that in Step 2, each composite proportion category would be divided by 3 (the total number of items).

• Activities
The nursing home score for this composite is produced by combining responses to third questions:
o Q33: "Are there enough organized activities for you to do on the weekends?"
o Q34: "Are there enough organized activities for you to do during the week?"

Respondents to the above questions can answer "yes", "no" or "sometimes" to each. The steps to calculate a nursing home's composite score for this domain are similar to Environment composite except that in Step 2, each composite proportion category would be divided by 2 (the total number of items).

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
For statistical significance for each composite or global item, we used 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)*:
Sampling Guidelines
Sampling Frame Elements
The following information must be included in the sample frame that a sponsor provides to the vendor. These data elements should come from the nursing home facility´s medical records of all current residents; most, but not all, data may be collected from the most current Minimum Data Set (MDS)3.0 available:
• Name
• Room number
• Legal guardian or other legal oversight
• Date of admission(note: Residents who return to the nursing home following any hospital discharge will not have their stay reset to zero when they return to the facility. (AHRQ will harmonize its specification on long stay residents with CMS)
• Comatose status (MDS 3.0 item B0100)
• Brief Interview for Mental Status (BIMS)score MDS 3.0 item C0500 or the staff assessment of mental status on cognitive skills for daily decision making" (MDS 3.0 C1000).
The following elements are also helpful in the interviewing process; if possible, these should be included in the sample frame as well:
• Patient unique nursing home identifier
• Gender (MDS 3.0 item A0800)
• Date of birth (MDS 3.0 item A0900)
Researchers have found the following elements to be potentially useful analytic variables:
• Race/Ethnicity (MDS 3.0 item A1000)
• Education (not available on MDS 3.0)
• Date of most recent MDS assessment
• Current payment source: Medicaid or Medicare

Sample Size
• The CAHPS Team's preliminary recommendation is to aim for a minimum of 50 completed interviews per facility. Based on our field test experiences, an initial sample size of 75 eligible residents may be needed to yield 50 completed interviews.
• Nursing homes that may not be able to achieve the recommended minimum of 50 completed interviews should attempt to interview all eligible residents.
• Nursing homes large enough to potentially yield more completed surveys than the recommended minimum should create a list of all eligible residents, randomize the list, then attempt to interview residents selecting in order from the randomized list until the targeted number of interviews is reached. Or, if they choose, they could interview additional residents after the target number of interviews is reached.
Eligible Population
A number of criteria define the population eligible to participate in the survey. To qualify as an eligible survey respondent:

- The resident must be 18 years or older.
- The resident must be living at the nursing home at the time of the initial visit by the interviewer.
- The resident must have been living at the nursing home for more than 100 days at the time of the initial visit by the interviewer (AHRQ will harmonize its day count specification with CMS).

If a resident has a legal guardian or other legal oversight, interviewers must have prior approval from the guardian or overseer before talking to the resident.

Excluded Populations

The only population excluded from the sample is residents who are comatose (as indicated on MDS); The nursing home may also choose to exclude from the sample residents who are severely impaired in mental status or skills for daily decision making (see MDS 3.0 items above). The CAHPS Team excluded this group from the field tests of this instrument. If these individuals are included, the sample size needs to be increased accordingly in order to yield the minimum number of completed interviews.

Response Rates

In its simplest form, the response rate is the total number of completed questionnaires divided by the total number of residents selected. For CAHPS analyses and reports, this rate is adjusted as shown in the following formula:

Number of completed questionnaires

Total number of residents selected – (deceased + ineligible)

In calculating the response rate, do not exclude residents who refused or who were unable to complete the questionnaire because of language barriers or cognitive difficulties.

Numerator Inclusions:

- Completed questionnaires. A questionnaire is considered complete if responses are available for at least 50 percent of the items that could be answered by all respondents (for a list of these key items, refer to Appendix: Determining Whether a Question Is Complete at https://www.cahps.ahrq.gov/content/products/NH/PROD_NH_Long-Stay_Prelim_Guidelines.htm). In addition, interviews in which residents who are unable to answer three questions in a row within the first six questions should be considered incomplete and thus excluded from the numerator.

Denominator Inclusions:

- Refusals. The resident (or guardian) refused to participate.
- Nonresponse. The resident is presumed to be eligible but did not complete the interview for some reason (for example, was unavailable at the time of the interview, was ill or cognitively unable to complete the survey, or had hearing problems or a language barrier).

Data Collection

The Long-Stay Resident Instrument must be administered in person by a trained interviewer. Sponsors should retain a third-party vendor with experience in in-person interviewing and interviewing an elderly/nursing home population.

Interviewers

The CAHPS Consortium recommends using professional interviewers to conduct the in-person interviews. Some studies have used graduate students, ombudsmen, or volunteers to conduct the interviews. These individuals should receive training in standardized interviewing techniques, particularly with an elderly/nursing home population. Individuals who provide care or services to the nursing home residents being surveyed should not be interviewers.

Privacy and Confidentiality

Privacy During the Interview

When possible, interviews should be conducted privately. However, interviewers might find it difficult to secure a private area for an interview. For example, a resident might not want to go to a private area, cannot be moved, or might prefer to be interviewed in his/her room with a roommate present. In these instances, interviewers should try to maintain as much privacy as possible (e.g., draw a curtain, allow the resident to point to responses on a show card [see Appendix: Showcards With Printed Response Options in CAHPS Nursing Home Survey – Long-Stay Resident Instrument (With Instructions) at website above], rather than giving an answer out loud). At no time should staff members, family, or friends be present during the interview. For example, if a staff person enters the room during the interview, the interviewer should stop the interview and wait until the staff person leaves.

Confidentiality of Responses

All information that could identify respondents must be kept confidential. The respondent's name must not

appear anywhere on the questionnaire; instead, unique identifiers should be placed on the cover page of the survey. In order to assure that respondents cannot be identified by their responses to the interview, our preliminary recommendation is that the vendor should not present summary data to the nursing home until 50 interviews per facility have been completed.

Minimum sample size:
The number of residents needed for each composite to reach a reliability of 0.70 (if the goal is public reporting for reliable comparison purposes) was calculated with the Spearman-Brown Prediction formula using the average number of respondents per nursing home. Based on the pilot test of the 2005 Resident survey, the following number of completes are needed to reach 0.70 reliability for the composites below:

Composite 1 Environment= 92.7
Composite 2: Care = 50.0
Composite 3: Communication & Respect = 55.9
Composite 4: Autonomy = 81.1
Composite 5: Activities = 29.5
So the minimum number of completes to be sufficient for all composites is 93. If necessary this data could be accumulated over time to achieve sufficient sample size. If the goal is to use survey data only for quality improvement purposes, a smaller number of completes may be used. (for more detail see Table 28 on page 88 of Harvard Final Report)

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested*)
Survey: Patient, Special or unique 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.*):
CAHPS® Nursing Home Survey: Long-Stay Resident Instrument

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
https://www.cahps.ahrq.gov/content/products/NH/NH_Long-Stay_Instrument.pdf

2a.29-31 Data dictionary/code table web page URL or attachment: Attachment   CODEBOOK FOR LONG STAY NURSING HOME RESIDENT final tues 5_11_10.doc

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   nurse aides

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: 2005 field test data from 13 nursing homes in four New England states (n= 439)

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

2b
C ☐
P ☐
M ☐
N ☐

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

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

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)*:
The Cronbach's alpha for each composite is:
Composite 1: Environment = 0.71
Composite 2: Care = 0.79
Composite 3: Communication and Respect = 0.86
Composite 4: Autonomy = 0.60
Composite 5: Activities = 0.60
Although a Cronbach's alpha of 0.70 is considered desirable, an alpha of 0.60 can be considered acceptable.

The Nursing Home reliability or inter-unit reliability (IUR) for each composite is:
Composite 1: Environment = 0.46
Composite 2: Care = 0.61
Composite 3: Communication and Respect = 0.58
Composite 4: Autonomy = 0.49
Composite 5: Activities = 0.72
Although the observed facility-level reliability of Composites 1, 2, 3 and 4 are not as high as we would like, it will be able to discriminate across nursing homes, given a sufficient number of respondents per facility.

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: 2005 field test data from 13 nursing homes in four New England states (n= 439)

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
We examined the correlation of each of the composites with the global ratings as a measure of criterion validity.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Correlation with Rating of Care from NH Staff
Composite 1: Environment = 0.55
Composite 2: Care = 0.63
Composite 3: Communication and Respect = 0.79
Composite 4: Autonomy = 0.20
Composite 5: Activities = 0.21

Correlation with Overall Rating of Nursing Home
Composite 1: Environment = 0.57
Composite 2: Care = 0.47
Composite 3: Communication and Respect = 0.55
Composite 4: Autonomy = 0.24
Composite 5: Activities = 0.28
Correlation with Would Recommend Nursing Home to Others
Composite 1: Environment = 0.45
Composite 2: Care = 0.33
Composite 3: Communication and Respect = 0.42
Composite 4: Autonomy = 0.20
Composite 5: Activities = 0.34

All five composites have statistically significant correlations (p <.001) with the three global measure, although the Autonomy composite has lower correlation (<0.30) than the other composites with all 3 global items; the Activities composite is lower than 0.30 on two global items.

For more detail see Table 27 b (Interview sample) on page 85 of Harvard Report.

2c
C☐
P☐
M☐
N☐

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.

2d. **Exclusions Justified**

2d.1 Summary of Evidence supporting exclusion(s):
Expert opinion was that a minimum of 30 days stay (without a planned discharge)in a nursing home was needed for residents to form a stable opinion of their experience. AHRQ will now harmonize with CMS meadure speciifcations for long stay nursing home resident to be defined as having a stay of more than 100 days, and if residents return to nursing home from any hospital discharge their day count will not be reset to zero. Excluding residents who were severely impaired in cognitive skills for daily decison making and may have interviewing problems was based on analyses of MDS data and nursing home researchers. Excluding persons in a coma is common sense.

2d.2 Citations for Evidence:
Expert opinion and sample frame development for field test

2d.3 Data/sample (description of data/sample and size): 2005 field test data from 13 nursing homes in four New England states (n= 439)

2d.4 Analytic Method (type analysis & rationale):
Residents were declared ineligible from the sampling frame in the pilot test if any of the following criteria were met (numbers refer to items in the MDS 2.0 record):
1. Those who had a guardian or other legal oversight (A9a = 1 or A9b = 1). Because of the tight schedule for the field test, it was not possible to take the time to gain consent from people outside the home. With another design or more time, this group of people would not have to be excluded.
2. Those whose last MDS evaluation indicated they were "severely impaired in cognitive skills for daily decision making" (B4 = 3).
3. Those who were in a coma (B1=1).
4. Those who had not been in the home for 30 days—or would not be by the time of data collection (AB1 < 30 days from interview date).
5. Those who had a discharge planned within 90 days (Q1c = 1 or 2).
6. those under age 18
During survey administration there were the following additional exclusions:
1. Non-English speaking (pilot survey only available in English)
2. unable to answer 3 questions in a row
3. conscious but unresponsive to interviewer

2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses):
see Table 2 through 4 on pages 25-28 of Harvard Final Report to see percentage breakdown of residents determined ineligible for several categories (overall 57% eligible); 31% of eligible sample were not able to be interviewed for a number of reasons.

2d
C ☐
P ☐
M ☐
N ☐
NA ☐

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

2e.1 Data/sample (description of data/sample and size): none conducted

2e.2 Analytic Method (type of risk adjustment, analysis, & rationale):
none conducted

2e.3 Testing Results (risk model performance metrics):
none conducted

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: During development of this long stay resident survey, the resources and activities were mostly concentrated on how to identify individuals who could respond, how best to stratify potential respondents and assess the mix of those who were and were not able to respond to a survey across nursing homes and the formidable sampling and surveying issues.

2e
C ☐
P ☐
M ☐
N ☐
NA ☐

2f. **Identification of Meaningful Differences in Performance**

2f
C ☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   13

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

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

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.

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: 2005 field test data from 13 nursing homes in four New England states (n= 439) **2f.2** Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*: For statistical significance we useds t-test comparing each nursing home meanto the mean of all the nursing home means for each composite **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)*: The mean and standard deviations (SD) for the composites are: Composite 1: Environment -- mean = 5.47 (0.97) Composite 2: Care-- mean = 6.88 (1.65) Composite 3: Communication and Respect- mean= 8.06 (1.99) Composite 4: Autonomy – mean= 2.80 (0.42) Composite 5: Activities – mean = 2.51 (0.66) Additional statistical detail on pages 83-84 of Harvard Final Report	P☐ M☐ N☐
2g. Comparability of Multiple Data Sources/Methods **2g.1** Data/sample *(description of data/sample and size)*: Alberta, Canada resident data not available at current time **2g.2** Analytic Method *(type of analysis & rationale)*: **2g.3** Testing Results *(e.g., correlation statistics, comparison of rankings)*:	**2g** C☐ P☐ M☐ N☐ NA☐
2h. Disparities in Care **2h.1** If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: not applicable **2h.2** If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: not applicable	**2h** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? Rationale:	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1** Current Use: In use **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*):	**3a** C☐ P☐ M☐ N☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   14

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

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

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

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

The Health Quality Council of Alberta, Canada, is using this survey for public reporting in aggregate -- see http://hqca.ca/index.php?id=130.

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*):
The Health Quality Council of Alberta, Canada, is using this survey for QI by providing site specific results back to nursing homes and comparing them to peers and norms. Also, this survey is included as one possible survey for nursing homes to use as part of Goal 7 (measuring Resident & Family Satisfaction) of the Advancing Excellence in America's Nursing Homes Campaign, of which more than 6400 U.S. nursing homes have joined (a home should pick three out of 8 possible goals).

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*): six focus groups in 3 states - four with nursing home residents and two with family members
Cognitive testing:
Round 1: 52 residents in 5 homes
Round 2: 15 residents in 3 homes
Round 3: 19 residents in 3 homes
Round 4: 27 residents in 3 homes
Round 5: 31 residents in 3 homes
Round 6: 16 residents in 2 homes
Round 7: 19 residents in 2 homes
For more detail, see Appendix A in Journal of Aging and Social Policy article on page 79

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.

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.

Development and Testing of Vignettes as a Potential Cognitive Screen

As previously discussed, throughout the cognitive interviewing process, the team was very concerned about
how to identify who could or could not participate in the NHCAHPS interview. Different kinds of screeners
were used, yet none tested specifically for the skills needed to answer the questionnaire.
The team reviewed the literature for instruments measuring short- and long-term memory, ability to
generalize, daily decision making, and recall. They examined and compared the Mini-Mental Status Exam,
EXIT25, Short Blessed, CLOX 1 and additional clock drawing tests, animal naming test, the Cognitive
Performance Scale of CPS, and others to determine which might best be suited for our research. The team
was not able to find a short screener that clearly addressed all of our needs. Many of the short
standardized assessments focus on temporal orientation; yet, orientation to place and people may be more
important for nursing home residents. Moreover, the literature does not provide good information on how
predictive temporal orientation may be of these other orientations.
One promising approach to cognitive screening identified through the literature review was a vignette
method. The team developed vignettes as a test of residents' abilities to generalize across positive and
negative experiences and to assign a numeric rating to abstract situations. The research team felt that the
vignettes should be about something with which most residents might need help. We developed a set of
three vignettes (Appendix C in JASP article) on rating of help with dressing with the same 0 to 10 rating
task using for QOC items. These vignettes were administered by interviews as part of the survey and were
never used as screeners (i.e., interviewers were never terminated because of how the respondents
answered). Instead, the vignette responses were scored after the interviews were completed and
compared to results from other cognitive measures and to the survey answers. Residents were evaluated
on their abilities to score the vignettes in a logical order, that is, given a better rating to a vignette in
which the depicted person always received help in dressing than to one with help most of the time, and a
higher rating to the vignette with help most of the time than to one getting no help in dressing. The
responses to the three vignettes were reviewed to ensure that the response pattern was as described.
Along with the set of vignettes, the interviewers administered the Short Blessed, a six-item test that covers
short-term memory, temporal orientation, and reasoning. In addition, the interviewers provided
assessments of their perceptions of residents' understanding during the interviews of the questions and of
the cognitive probes. MDS data were also collected for each resident in order to compare CPS scores
against these other measures.
The vignette error score was predictive of the percentage of questions that had missing responses such
that, compared to others, respondents who had higher vignette error rates also tended to have lower
percentages of survey questions answered. For example, respondents with three errors on the vignettes
answered only 77% of questions in the survey, on average, compared with respondents with no errors on the

vignettes who answered 95% of questions in the survey. The team also looked at the mean number of questions answered with in-range responses (i.e., resident responded on the 0 to 10 scale). Within CPS levels, the vignette score generally provided additional information to help distinguish those better able to answer from those less able to do so. For example, within a CPS score of 1, persons with a zero vignette errors gave 100% in-range responses on a average, while persons with three errors gave only 81% in-range responses on average. We also found mean vignette error scores were more consistent with interviewer observations that CPS scores. Residents judged by interviewers as not understanding the probes at all had higher mean vignette error scores (i.e., made more errors in scaling the vignettes) compared with those judged by interviewers as understanding probes "sometimes." The mean CPS score was 2.0 for both groups and did not differentiate.

In summary, the vignettes used in the cognitive interviewing appear to enhance the ability of the CPS to identify individuals who respond to higher percentages of questions. They also showed a more consistent relationship with interviewer confidence ratings than the CPS alone.

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

3b.1 NQF # and Title of similar or related measures:
There are similar CAHPS survey measures but for different types or settings of care (Hospital CAHPS, Clinician and Group CAHPS, Home health CAHPS). Separate measures are being submitted to NQF for family members of nursing home residents and for short-stay nursing home residents.

(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?
yes, the measure specifications of this CAHPS nursing home resident instrument is harmonized with other CAHPS survey measure specifications.

| 3b |
| C ☐ |
| P ☐ |
| M ☐ |
| N ☐ |
| NA ☐ |

> **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.

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:
not applicable

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:
There is no similar measure for the same target population. This is the only measure for long stay nursing home resident experience.

| 3c |
| C ☐ |
| P ☐ |
| M ☐ |
| N ☐ |
| NA ☐ |

> **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).

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

| 3 |

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

| 3 |
| C ☐ |
| P ☐ |
| M ☐ |
| N ☐ |

---

4. FEASIBILITY

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

| Eval Rating |

4a. Data Generated as a Byproduct of Care Processes

4a.1-2 How are the data elements that are needed to compute measure scores generated?
Survey

| 4a |
| C ☐ |
| P ☐ |
| M ☐ |
| N ☐ |

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

4b. Electronic Sources

| 4b |

> **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   19

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

4b.2 If not, specify the near-term path to achieve electronic capture by most providers.
this is an in-person survey instrument so electronic capture is not considered; only MDS items for sampling frame may be electronically available

C☐	
P☐	
M☐	
N☐	

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.

4c
C☐
P☐
M☐
N☐
NA☐

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

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

4d.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., if the interviewers do not ask each of the questionnaire items as worded on the survey, or the interviewer did not assure privacy of resident in the interview). 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 Family Member 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☐

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

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
Lessons learned:
The Harvard Field Test Report (see attached) describes the results of a field test that was conducted, as part of the survey development process, to learn more about how samples of potential respondents would be identified, how best to work with nursing homes to identify potential respondents, how best to conduct surveys, and about the performance of the draft survey. Previous work by the CAHPS consortium determined that the most feasible and accurate method of surveying nursing home residents most likely would differ for short and long term residents. Thus, the pilot study included two distinct activities - in-person interviewing of long term nursing home residents and a mail survey of recently discharged residents.

Protocol: An important part of the survey protocol was how interviewers were to decide who was able to be interviewed. In addition to the survey questions, the interview had a series of three vignettes about hypothetical residents' experiences in nursing homes that were thought to be a good predictor of the ability to answer the survey questions. The Short Blessed (a frequently used test of cognitive ability) was administered at the end of the survey. Interviewers used neither of these to screen out respondents. Rather, they tried to ask every assigned respondent all the survey questions. If the respondent could not provide a meaningful answer to any three questions in a row, the interview was terminated.

PRETEST: On May 26, 2005, the protocol and survey were pretested. We learned many things from the pretest. Using a single person as "site coordinator" to manage and control the sample worked well. Finding private locations to do interviews was a challenge. Showing the respondents the response options on a show card was helpful to both the respondent and the interviewer during the interview process. We also found that when talking with respondents who had cognitive difficulties, it was necessary to add an "unresponsive" code - to be used when the respondent was conscious but totally disoriented or unresponsive to the interviewer. Based on what was learned during the pretest, we also changed the

4e
C☐
P☐
M☐
N☐

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

wording of some questions and simplified the informed consent script page.

Field Test Results:
Sampling: We asked each nursing home to provide 19 items from all of their current residents' Minimum Data Set (MDS) data. This information included basic demographics, items we needed for sampling, and items needed to create a Cognitive Performance Scale (CPS) score. Almost all the homes had the information needed in electronic form, but the majority lacked either the data processing expertise, or the staff time, to produce selected data from their files. For future studies using this protocol, we feel the best way to do the sampling is to collect the data that is necessary to define the sample from the nursing home and then have the project staff actually process the information to select the sample.

Eligibility: The 12 nursing homes sent a total of 1347 names of current residents; 57% were eligible for the long-term stay survey. If we include those residents who had guardians or other legal overseers and who were not ineligible for other reasons and those who probably could be interviewed in another language, that number rises to 67%. At the individual nursing home level, the rates of eligibility range from 36.1% to 93.0%. The presence of specialized Alzheimer's or psychiatric treatment units and the percentage of short term beds are the two factors that seem to most influence this rate.

Data Collection Results: Of the 870 residents who were believed to be eligible based on analysis of the record data provided, 103 were found or estimated to be ineligible and another 169 were not contacted because they were not needed to meet targeted sample goals. Thus, there were 618 residents whom interviewers attempted to interview who were part of the study population. Of those, interviews were completed with 424 residents, which is 69% of the eligible sample that interviewers attempted to interview. The most common reason for nonresponse was that eligible respondents were cognitively unable to answer survey questions; 39% of nonrespondents were unable to answer 3 questions in a row, 22% could not be roused to answer any questions at all. Thus, close to 20% of the total eligible sample and 61% of the nonrespondents were not able to do an interview. Most of the other nonresponse was due to hearing problems, not feeling well, and not being willing to be interviewed. However, all together, those reasons accounted for less than 12% of the total sample not being interviewed. We conclude that most of residents who are physically and cognitively able to be interviewed are willing to do so. The protocol called for interviewers to go back to all respondents who initially were busy, ill, unresponsive, or who had refused. The idea was that finding a "better time" would lead to getting interviews. For refusals, a different interviewer made the second interview attempt; 95% of all those interviewed were interviewed on the first or second contact with an interviewer. Contacting nonrespondents a third time to try to complete an interview was not productive.

Screening for ability to respond: We think all eligible residents should be approached and that interviewers should not rely on medical records or staff members to determine appropriateness for interviewing. By only eliminating the most severely impaired (those with a CPS score of 5 or 6), we were able to interview some respondents with moderately high impairment (and CPS scores) who might be eliminated in other protocols. Interviewers would prefer not to use a screener for cognitive ability unless it is highly predictive. We feel that the best way to screen for ability to complete the interview is to actually attempt to do the interview. If a respondent is unable to answer 3 questions in a row, then the interview should be stopped. (This is similar to procedure to be used when MDS 3.0 is implemented)

Data Collection Process: It was not easy to find a private place to administer the interview. Even for those interviews that were done with other people around, however, interviewers felt that it rarely interfered with the survey process. Part of this could be because of the use of show cards. As expected, many respondents who were interviewed had physical and intellectual impairments. Interviewers felt that only about 66% of residents were always able to understand the survey questions.

Length of Interview Schedule: The length of the interviews worked well. In about 83% of the cases, the survey itself (not including vignettes or the Short Blessed) took 20 minutes or less to complete. There were only 15 of the 424 interviews that took more than 30 minutes to complete and most of these took that long because the respondents liked to talk and it was sometimes hard to keep them focused on the interview.

Feedback from Nursing Home Administrators: Almost all of the administrators felt the sampling process went well. Administrators said it took an average of 8 hours to access and compile the data we requested

of them. This number depended on how the records are kept at the home, the person's familiarity with the computer systems, and whether CSR sent staff to the home to collect the information or it was sent to us. Since our original data request was for both the current and discharged residents, some amount of time (and some problems) may be the result of getting the data for residents who are no longer there. When asked, most said they could have created lists for us of residents who met certain sampling criteria, but considering the problems of getting simple census data from these sites, we think it would be difficult for the homes to do the sampling required correctly. All of the nursing homes thought the actual interviewing process went well and were pleased with the self-sufficiency of the interviewing team. On the whole, there were no disruptions or difficulties.

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 direct costs (excluding travel and overhead) for the 2005 pilot test in 12 nursing homes was $24000 or about $57 per completed interview. For more detail, see pages 6 and 36-39 of the Field Test report.

4e.3 Evidence for costs:
The 2005 pilot test in 12 nursing homes. Additional cost information is available from Alberta, Canada (using slightly modified CAHPS survey with trained graduate students instead of professional interviewers). The Ohio Department of Aging (using a similar in-person survey) spent $980,000 for interviewing 32,561 residents of both nursing homes (n=960) and assisted living (n=560) or $30.10 per completed resident interview in sample.

4e.4 Business case documentation: Assessing resident satisfaction is the first step in making changes or improvements in the quality of the care and quality of life in the nursing home. A survey allows residents the chance to report their experience with care and daily life in the nursing home. Although it is less expensive to conduct a mail survey with family members, family members views often differ from those of the residents. Ohio has used an approach of alternating years for conducting an in-person resident survey and a mail survey with family members.

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.

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? Rationale:	4 C ☐ P ☐ M ☐ N ☐

RECOMMENDATION	
(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.	Time-limited ☐
Steering Committee: Do you recommend for endorsement? Comments:	Y ☐ N ☐ A ☐

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
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: 2006
Ad.7 Month and Year of most recent revision:
Ad.8 What is your frequency for review/update of this measure? CAHPS team is reviewing mixed reporting composites and environment composite question design.
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.pdf

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

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 that has the greatest effect on improving the specified desired outcome(s).
  - o <u>Structure</u> – 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.
  - o <u>Patient experience</u> – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - o <u>Access</u> – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - o <u>Efficiency</u> – 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.

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.

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.