NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Nursing Homes 2010

Measure Number/Title: NH-027-10: Consumer Assessment of Health Providers and Systems (CAHPS®) Nursing Home Survey: Discharged Resident Instrument

Description: 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.

Numerator Statement: 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)

Denominator Statement: The denominator is the total number of surveys for respondents that meet CAHPS completion standard and any applicable screener.

Level of Analysis: Facility/Agency

Data Source: Survey: Patient, special or unique data

Measure developer: Agency for Healthcare Research and Quality

Type of Endorsement (full or time-limited): Full

Attachments: JASP_resident NHCAHPS; Integrated Nursing Home CAHPS Report; Codebook for Long Stay Nursing Home Residents; Nursing Home Final Report_17_Sept08

NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

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

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

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

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

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

(for NQF staff use) **NQF Review #:** NH-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 both public reporting and 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.1Testing:** 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 Ratin g
(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**☐

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 ☐

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step 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, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rating
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** 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 are under age 18, comatose, severely impaired in cognitive skills for daily decisionmaking, those who cannot answer 3 questions in a row; conscious but unresponsive to interviewer and unable to speak English for survey. 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 are discharged to a hospital with return anticipated will not have the 100 days count reset to zero when they return to the facility.

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

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
1. Residents who had a guardian or other legal oversight (A9a = 1 or A9b = 1) or under age 18.
2. Residents whose last MDS evaluation indicated they were "severely impaired in cognitive skills for daily decision making" (B4 = 3).
3. Residents who were in a coma (B1=1).
4. Residents 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. Residents who had a discharge planned within 90 days (Q1c = 1 or 2).
During survey administration there were the following additional exclusions determined by trained interviewers:
1. Non-English speaking (pilot survey only available in English)
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***):**

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

The nursing home score for this composite is produced by combining responses to eight questions:

- Q1: "What number would you use to rate the food here at the nursing home?"
- Q3: "When you eat in the dining room, what number would you use to rate how much you enjoy mealtimes?"
- Q4: "What number would you use to rate how comfortable the temperature is in the nursing home?"
- Q5: "What number would you use to rate how clean the nursing home is?"
- Q6: "What number would you use to describe how safe and secure you feel in the nursing home?"
- Q18: "Is the area around your room quiet at night?"
- 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)
- 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:
- Q8: "What number would you use to rate how well the medicine worked to help with aches or pain?"
- Q9: "What number would you use to rate how well the staff help you when you have pain?"
- Q10: "What number would you use to rate how quickly the staff come when you call for help?"
- Q12: "What number would you use to rate how gentle the staff are when they're helping you?"
- 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:
- Q13: "What number would you use to rate how respectful the staff are to you?"
- Q14: "What number would you use to rate how well the staff listen to you?"
- 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:
- Q30: "Can you choose what time you go to bed?"
- Q31: "Can you choose what clothes you wear?"
- 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:
- Q33: "Are there enough organized activities for you to do on the weekends?"
- 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. Data should be collected from the most current Minimum Data Set (MDS) available:
- Name
- Room number
- Legal guardian or other legal oversight (MDS items A9a, A9b)
- Date of admission (MDS item AB1)
- Comatose status (MDS item B1)
- Whether discharge was planned (MDS item Q1c)
- Cognitive skills for daily decision making (MDS item B4)

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 item AA2)
- Date of birth (MDS item AA3)

Researchers have found the following elements to be potentially useful analytic variables:
- Race/Ethnicity (MDS item AA4)
- Education (MDS item AB7)
- Date of most recent MDS assessment (MDS item A3)
- Current payment source: Medicaid or Medicare (MDS item A7a, A7b)

The Cognitive Performance Scale (CPS) score is also a useful analytic variable. The following elements are needed to create the CPS score:
- Short-term memory problems (MDS item B2a)
- Cognitive skills for daily decision making (MDS item B4)
- Making self understood (MDS item C4)
- Eating self-performance (MDS item G1hA)

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 at least 30 days at the time of the initial visit by the interviewer.
- The resident must have no discharge planned within 90 days (as indicated on MDS; MDS item Q1c=0 or 3).

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; MDS item B1=1).

The nursing home may also choose to exclude from the sample residents who are severely impaired in skills for daily decision making (MDS item B4=3) because this would exclude those with the greatest cognitive impairment (CPS levels 5 and 6). 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

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.

an item or composite (IUR = (F-1)/F). IUR can be interpreted as the fraction of the variation among facility scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or composite to discriminate across facilities is greater. An IUR > 0.70 is considered to indicate a high level of discriminant ability for an item or composite. As the IUR gets smaller, you need a larger sample in order to reliably discriminate across facilities.

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

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.

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.

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. 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 if any of the following criteria were met (numbers refer to items in the MDS 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 n 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

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☐

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

2g. Comparability of Multiple Data Sources/Methods

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☐

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

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: 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☐

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.

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☐

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). If not used for QI, 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☐

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☐

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☐

4b. Electronic Sources

4b

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

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

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

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

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

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) estimated costs of $980,000 for 35,000 residents of both nursing homes (n=960) and assisted living (n=560) or approximately $28 per resident in sample (plan to obtain information on number of completes soon to get equivalent calculation)

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*)**:** 07/13/2010

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

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
  - Intermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
  - Process – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - Structure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
  - Patient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
  - Access – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
  - Efficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

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

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

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

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

ILL Document Delivery

REG-14122759

MDUAGE

NLM -- W1 JO534BK (Gen)

Agency for Healthcare Research and Quality
Information Resources Center
540 Gaither Road Room 1000
Rockville, MD 20850-6649

ATTN:
PHONE: 301-427-1280
FAX: 301-427-1286
E-MAIL: IRCPatronRequests@ahrq.hhs.gov

SUBMITTED: 2007-05-17 08:16:04
PRINTED: 2007-05-18 09:14:31
REQUEST NO.: REG-14122759
SENT VIA: DOCLINE
DOCLINE NO: 22549041

REG **Copy** **Journal**

TITLE:	JOURNAL OF AGING & SOCIAL POLICY
PUBLISHER/PLACE:	Haworth Press, [Binghamton, NY] :
VOLUME/ISSUE/PAGES:	2007;19(2):63-82 63-82
DATE:	2007
AUTHOR OF ARTICLE:	Sangl J;Buchanan J;Cosenza C;Bernard S;Keller S ;Mitchell N;B
TITLE OF ARTICLE:	The development of a CAHPS instrument for Nursing
ISSN:	0895-9420
OTHER NUMBERS/LETTERS:	
	NLM Unique ID: 8914669
	PubMed ID: 17409047
SOURCE:	PubMed
MAX COST:	$15.00
COPYRIGHT COMP.:	Guidelines
CALL NUMBER:	W1 JO534BK (Gen)
NOTES:	**Federal Library; FreeShare Library**
REQUESTER INFO:	McCullough - AIA (PC11)
DELIVERY:	E-mail: IRCPatronRequests@ahrq.hhs.gov
REPLY:	Mail:

KEEP THIS RECEIPT TO RECONCILE WITH BILLING STATEMENT

For problems or questions, contact NLM at http://wwwcf.nlm.nih.gov/ill/ill_web_form.cfm or phone 301-496-5511.

Include LIBID and request number.

NOTE: THIS MATERIAL MAY BE PROTECTED BY COPYRIGHT LAW (TITLE 17, U.S. CODE)

The Development of a CAHPS® Instrument for Nursing Home Residents (NHCAHPS)

Judith Sangl, PhD
Agency for Healthcare Research, Rockville, MD

Joan Buchanan, PhD
Harvard Medical School

Carol Cosenza, MSW
University of Massachusetts Boston

Shula Bernard, PhD
Research Triangle Institute, Washington, DC

Susan Keller, PhD
American Institutes for Research, Chapel Hill, NC

Nancy Mitchell, BA
Research Triangle Institute

Julie Brown, BA
The Rand Corporation, Santa Monica, CA

Nicholas Castle, PhD
University of Pittsburgh

Edward Sekscenski, MPH
Deborah Larwood, DM
Centers for Medicare and Medicaid Services

Judith Sangl, PhD, is Health Scientist Administrator, Agency for Healthcare Research and Quality.

Joan Buchanan, PhD, is Lecturer, in Health Care Policy, Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115.

Journal of Aging & Social Policy, Vol. 19(2) 2007
Available online at http://jasp.haworthpress.com
© 2007 by The Haworth Press, Inc. All rights reserved.
doi:10.1300/J031v19n02_04

Material may be protected by copyright law (Title 17, U.S. Code)

ABSTRACT. We report on a federal initiative to develop a CAHPS® (The Consumer Assessment of Healthcare Providers and Systems) survey to measure residents' experiences with quality-of-care and quality-of-life in nursing homes (known as NHCAHPS). We focus on how we created and tested questions for inclusion in the instrument and tested a possible cognitive screener to determine which residents could participate in a NHCAHPS interview. The major lessons learned were: (1) In contrast to other CAHPS surveys, ratings were more useful than reports because of the difficulty that residents had with summarizing over time and people; (2) consistent with other CAHPS surveys, the 0 to 10 response scale appeared to work well with nursing home residents for many of the quality-of-care questions; however, a different response scale was needed for many of the quality-of-life items; and (3) in contrast with typical survey methodology and other CAHPS surveys where explicit time reference periods are used, a non-specific present reference period in questions seemed to work best. doi:10.1300/J031v19n02_04
[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com> © 2007 by The Haworth Press, Inc. All rights reserved.]

Carol Cosenza, MSW, is Project Manager, Center for Survey Research, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125.

Shula Bernard, PhD, is Director, of the Program on Health Care Quality and Outcomes at Research Triangle Institute, 1615 M Street, NW, Suite 740, Washington, DC 20036.

Susan Keller, PhD, is Principal Research Scientist, American Institutes for Research, 1506 E Franklin Street, Suite 200, Chapel Hill, NC 27514.

Nancy Mitchell, BA, is Policy Analyst, Research Triangle Institute, 1615 M Street, NW, Suite 740, Washington, DC 20036.

Julie Brown, BA, is Director, Survey Research Group, The Rand Corporation, 1776 Main Street, PO Box 2138, Santa Monica, CA 90407-2138.

Nicholas Castle, PhD, is Assistant Professor, Department of Health Policy and Management, at the University of Pittsburgh, 0649 Crabtree Hall, 130 DeSoto Street, Pittsburgh, PA 15261.

Edward Sekscenski, MPH, is CAHPS Team Leader, and Deborah Larwood, DM, is Health Insurance Specialist, Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Baltimore, MD 21228.

Address correspondence to: Dr. Judith Sangl, Health Scientist Administrator, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850 (E-mail: jsangl@ahrq.gov).

Portions of this paper were previously presented by Carol Cosenza and Floyd J. Fowler as "Building Experiments into Cognitive Testing," Proceedings of the 5th QUEST Workshop, Centraal Bureau voor de Statisteik, (Heerlen, Netherlands), April 19-21, 2005.

KEYWORDS. Patient experience
quality-of-life

INTRODU

The Consumer Assessment of He
(CAHPS®) is a family of survey ins
report people's experiences obtaining
on a federal initiative to develop a
to measure nursing home residents' ex
"NHCAHPS"). Federal and state gove
billion of total nursing home costs
alHealthExpendData/02_NationalHeal
federal agency responsible for nursing
for Medicare & Medicaid Services (CM
perspective on the quality of the care
this NHCAHPS initiative. NHCAHP
with the Agency for Healthcare Resea
CAHPS® consortium of Harvard Med
tion, Research Triangle Institute Inter
tutes for Research (AIR). (AIR took th
third member of the CAHPS consorti
phase of CAHPS.)

This article focuses on three areas
opment project: (1) Creating and testi
(2) testing the use of vignettes as poss
mine which residents could participat
(3) designing the field test. We end w
from the work to date and implicatio
similar settings.

BACKGRO

The consumer perspective on healt
a valid measure of care quality and as a
measures (IOM, 2001). The nursing hon
of-care (QOC) is commingled with qua
home residents live in the same place
Because discrepancies have been not

Material may be protected by copyright law (Title 17, U.S. Code)

federal initiative to develop a CAHPS® Healthcare Providers and Systems) surriences with quality-of-care and quality-n as NHCAHPS). We focus on how we inclusion in the instrument and tested a determine which residents could partici-. The major lessons learned were: (1) In ys, ratings were more useful than reports sidents had with summarizing over time other CAHPS surveys, the 0 to 10 re- well with nursing home residents for uestions; however, a different response he quality-of-life items; and (3) in con-dology and other CAHPS surveys where are used, a non-specific present refer-l to work best. doi:10.1300/J031v19n02_04

from The Haworth Document Delivery Ser-ddress: <docdelivery@haworthpress.com> ss.com> © 2007 by The Haworth Press, Inc.

iger, Center for Survey Research, University Blvd., Boston, MA 02125.
e Program on Health Care Quality and Out-5 M Street, NW, Suite 740, Washington, DC

arch Scientist, American Institutes for Re-, Chapel Hill, NC 27514.
, Research Triangle Institute, 1615 M Street,

esearch Group, The Rand Corporation, 1776 , CA 90407-2138.
rofessor, Department of Health Policy and rgh, 0649 Crabtree Hall, 130 DeSoto Street,

Team Leader, and Deborah Larwood, DM, is ledicare & Medicaid Services, 7500 Security

ith Sangl, Health Scientist Administrator, ity, 540 Gaither Road, Rockville, MD 20850

y presented by Carol Cosenza and Floyd J. Cognitive Testing," Proceedings of the 5th de Statisteik, (Heerlen, Netherlands), April

KEYWORDS. Patient experience, satisfaction survey, quality-of-care, quality-of-life

INTRODUCTION

The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) is a family of survey instruments designed to capture and report people's experiences obtaining medical care. This article reports on a federal initiative to develop a CAHPS® Nursing Home Survey to measure nursing home residents' experiences (hereafter referred to as "NHCAHPS"). Federal and state governments pay for 67% of the $110.8 billion of total nursing home costs (http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp). As the federal agency responsible for nursing home quality oversight, Centers for Medicare & Medicaid Services (CMS) is interested in the consumers' perspective on the quality of the care they receive and has supported this NHCAHPS initiative. NHCAHPS is developed in collaboration with the Agency for Healthcare Research and Quality (AHRQ) and the CAHPS® consortium of Harvard Medical School, The RAND Corporation, Research Triangle Institute International, and the American Institutes for Research (AIR). (AIR took the place of RTI International as the third member of the CAHPS consortium of grantees during the second phase of CAHPS.)

This article focuses on three areas of the resident NHCAHPS development project: (1) Creating and testing questions for the instrument; (2) testing the use of vignettes as possible cognitive screeners to determine which residents could participate in a NHCAHPS interview; and (3) designing the field test. We end with a summary of lessons learned from the work to date and implications for future research in this and similar settings.

BACKGROUND

The consumer perspective on health care has become accepted as a valid measure of care quality and as a complement to traditional clinical measures (IOM, 2001). The nursing home setting is unique in that quality-of-care (QOC) is commingled with quality-of-life (QOL), since nursing home residents live in the same place where they receive their care. Because discrepancies have been noted between proxy respondents

Material may be protected by copyright law (Title 17, U.S. Code)

(e.g., families or staff) and residents (Kane et al., 2000), the voice of the resident, particularly for QOL, is considered to be the "gold standard" (Kane et al., 2003). This preference for the resident self-report presumes the ability to report. However, the extent of cognitive impairment among nursing home residents can be a major impediment to the ability to obtain valid and reliable survey responses.

The goal of any survey, including NHCAHPS, is to maximize survey participation while not compromising the validity of survey responses. In many studies, the nursing home staff is asked to help select residents capable of participating in research studies (Simmons et al., 1997). It is common for staff not to recommend a large proportion of nursing home residents for interviews because the staff does not deem them fit or capable for participation; these residents are often classified by staff as "confused," "disoriented," or as "having difficulty communicating" and believed to have inadequate cognitive ability (Allen, Hogg, & Peace, 1992; Myers & MacDonald, 1996; Phillips et al., 1993). However, other studies have shown that cognitively impaired nursing home residents who are able to communicate verbally are able to report their satisfaction with the care they receive as well as their care preferences (Sansone et al., 1998; Mozley et al., 1999). One study found that 30% of residents successfully interviewed would have been excluded if selection for participation had been determined by staff perception and Minimum Data Set (MDS) scores instead of structured interviewer attempts to assess resident responsiveness (van Maris et al., 1996). Such interviewer training protocols have been developed as one method to maximize survey participation by engaging residents a certain number of times (often three) at various times of day before giving up on their participation (van Maris et al., 1996). While these methods may increase response rates, they do not give the information needed to determine the validity of the response. Our research follows on this work and attempts to ascertain a simple, unbiased way to figure out who can, and cannot, participate in an interview.

Development Process

The development of the NHCAHPS resident instrument has been a multi-phase process (Appendix A). In Phase 1, 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. At the end of Phase 1, CMS, AHRQ, and the CAHPS

team concluded that it was feasible t
ences in the nursing home from many
by conducting in-person surveys.

Phase 2 consisted of question deve
low the basic CAHPS principle of t
which the consumer is the best or o
Kojetin et al., 1999).

CMS initially requested that the ins
emphasizing care processes, especiall
the nursing home resident's perspecti
to avoid duplication of effort. CMS w
velopment of a survey by Dr. Rosalie
on residents' QOL (Kane et al., 200
CMS directed AHRQ and the CAHI
instrument with a subset of questions f
merged instrument would create a mo
uating both the QOC in the nursing h
QOL of nursing home residents. Such
existing resident satisfaction instrum

Domain Selection

To identify aspects of quality of ca
home residents, the team did the follo
review, including a review of public
al., 2000), (2) conducted phone interv
and (3) received advice from a Metho

Through these methods, several
These included (1) Nursing and med
rity; (3) getting needed help from nurs
communication with staff; (5) food; (
ness; (8) noise; (9) activities; and (10

Focus Groups

Focus groups help researchers learn
lation and appropriate vocabulary f
Three research organizations conduc
nursing home residents and two with
verify that the domains of care found
nursing home residents and to identify

Material may be protected by copyright law (Title 17, U.S. Code)

ts (Kane et al., 2000), the voice of the considered to be the "gold standard" nce for the resident self-report pre- er, the extent of cognitive impairment ı be a major impediment to the ability responses.

ıg NHCAHPS, is to maximize survey sing the validity of survey responses. staff is asked to help select residents h studies (Simmons et al., 1997). It is ıd a large proportion of nursing home ıe staff does not deem them fit or ca- dents are often classified by staff as ıaving difficulty communicating" and ıitive ability (Allen, Hogg, & Peace, Phillips et al., 1993). However, other ely impaired nursing home residents bally are able to report their satisfac- ell as their care preferences (Sansone One study found that 30% of residents ıve been excluded if selection for par- staff perception and Minimum Data ctured interviewer attempts to assess s et al., 1996). Such interviewer train- d as one method to maximize survey its a certain number of times (often fore giving up on their participation ıese methods may increase response tion needed to determine the validity llows on this work and attempts to to figure out who can, and cannot,

AHPS resident instrument has been a). In Phase 1, CMS requested AHRQ ıte the methodological challenges of ıome residents. This phase examined s, data collection methods, and possi- ıase 1, 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.

Phase 2 consisted of question development. The team decided to follow the basic CAHPS principle of the survey consisting of items for which the consumer is the best or only source of information (Harris-Kojetin et al., 1999).

CMS initially requested that the instrument focus exclusively on QOC, emphasizing care processes, especially interpersonal aspects of care from the nursing home resident's perspective. In part, this decision was made to avoid duplication of effort. CMS was concurrently sponsoring the development of a survey by Dr. Rosalie Kane and colleagues that focused on residents' QOL (Kane et al., 2004). However, in the fall of 2003, CMS directed AHRQ and the CAHPS team to merge the NHCAHPS instrument with a subset of questions from the QOL instrument. This new merged instrument would create a more comprehensive method of evaluating both the QOC in the nursing home and the effect of this care on QOL of nursing home residents. Such an instrument is similar to many existing resident satisfaction instruments that combine both aspects.

Domain Selection

To identify aspects of quality of care that were meaningful to nursing home residents, the team did the following: (1) It conducted a literature review, including a review of publicly available surveys (Edwards et al., 2000), (2) conducted phone interviews with long-term care experts, and (3) received advice from a Methodological Expert Group meeting.

Through these methods, several domains of interest were found. These included (1) Nursing and medical services; (2) safety and security; (3) getting needed help from nursing home staff; (4) interaction and communication with staff; (5) food; (6) dignity and respect; (7) cleanliness; (8) noise; (9) activities; and (10) overall ratings of care.

Focus Groups

Focus groups help researchers learn about issues important to a population and appropriate vocabulary for those issues (Fowler, 1995). Three research organizations conducted six focus groups (four with nursing home residents and two with family members) in three states to verify that the domains of care found in the literature were important to nursing home residents and to identify any other domains of interest. In

Material may be protected by copyright law (Title 17, U.S. Code)

addition, the family member focus groups were conducted to compare domains of interest with the resident groups and to learn about whether family members would be able to be proxy informers about the nursing home care experience.

Nursing homes near the research organizations were recruited by letters and personal contact with researchers. At each home, residents listed by staff as being able to consent to medical procedures (a human subjects protocol requirement) were approached and recruited by researchers. Family members were recruited either by the nursing home staff or through a local vendor. The focus group transcripts were analyzed using a transcript-based analysis method, used in other CAHPS® efforts. This method entails reviewing transcripts of focus group audiotapes to sort the transcript by grouping all text on a specific topic, identifying common themes or issues that emerge in the group discussion, looking at the frequency of specific comments, and identifying themes or issues that were unique to individual participants or groups.

Resident focus groups indicated that issues of greatest concern were cleanliness of the facility, noise, food, training, and 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 to our participants. Likewise, many residents reported the CAHPS domain, "communication with doctors," as being irrelevant to their QOC because they did not see doctors as often as other staff. Since CAHPS was originally created for use in ambulatory medical 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 basic day-to-day activities. We also learned from the family groups that they may not be 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 respondents' answers accurately reflected what they have to say on the topic. Cognitive interviewing is often done using small numbers of individuals (usually

5 to 15 respondents) in several iterati
veys are reviewed, the questions ar
and additional interviews are comp
qualitative method does not result
can be analyzed, it helps researcher
and answer questions and is invalu
stood survey questions, thus produc
followed a semi-structured protoco
tions and a set of scripted cognitiv
protocol called for interviewers to a
tain answers to one or a short serie
cognitive probes. The team used pr
total of seven rounds of cognitive
nursing homes near the research org
and personal contact with researche

Round 1. Nursing homes provide
residents. They were asked to incl
answer our interviewer-administere
probably have difficulty but could d
daily decision making or who cann
stood), and some who were unlikely
(those with memory problems, mor
sion making, or who often have di
stood). Interviewing teams talked to
the study and the interview process,
nitive screener. The screener consist
ety of other screeners intended to tes
If a respondent answered six or mor
eligible to be interviewed in this ro
failed this cognitive screener.

The goal of the first round was to
concepts in the draft survey. A particu
respondents could handle a four-cat
ally, sometimes, never) or if a two-c
worked better. After testing, we fou
sponse task, one of the core response
very difficult for nursing home resi
choice ("yes" or "no") as an alternati
ing out any sense of frequency in a qu
did you get help washing your face
question ambiguous rather than ma

Material may be protected by copyright law (Title 17, U.S. Code)

roups were conducted to compare groups and to learn about whether proxy informers about the nursing

rganizations were recruited by let-ers. At each home, residents listed dical procedures (a human subjects ied and recruited by researchers. ier by the nursing home staff or ip transcripts were analyzed using ed in other CAHPS® efforts. This of focus group audiotapes to sort a specific topic, identifying com-ie group discussion, looking at the identifying themes or issues that s or groups.

at issues of greatest concern were od, training, and competency of aff, and receiving correct medica-erature, such as "safety," were not cipants. Likewise, many residents unication with doctors," as being y did not see doctors as often as ally created for use in ambulatory some domains are inappropriate what was learned in the resident groups. The main concerns of the ability of activities, and adequacy s about medical care were much sic day-to-day activities. We also they may not be knowledgeable ng home residents.

ment

cus groups and literature review, nducted a series of cognitive inter-items were understood in a con-to learn whether the respondents' have to say on the topic. Cognitive ll numbers of individuals (usually

5 to 15 respondents) in several iterative rounds. After each round, the surveys are reviewed, the questions and protocol are modified as needed, and additional interviews are completed (Willis, 2004). Although this qualitative method does not result in actual numbers or statistics that can be analyzed, it helps researchers learn how respondents understand and answer questions and is invaluable in creating universally understood survey questions, thus producing better quality data. 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 (Appendix A). Again, nursing homes near the research organizations were recruited by letters and personal contact with researchers.

Round 1. 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 severe 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 core 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

Material may be protected by copyright law (Title 17, U.S. Code)

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 asked respondents how they decided 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 task easier. In many of the 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., "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 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.

Material may be protected by copyright law (Title 17, U.S. Code)

whether one *always* got help or *ever*
vas essential to making the question
eaningful.
in general, paid almost no attention
s. Yet, perhaps the most important
gnitive interviews was that summa-
a major challenge for respondents.
y decided on their answers, we found
o simplify the cognitive task by fo-
ngle event, thereby making the task
, the events asked about occur fre-
as events very much. For example,
ast week that eating or going to the
or respondents to synthesize. They
ow often these very common occur-
y of those times they had problems.
ound of cognitive interviewing, the
ating on question content, they first
of question most nursing home res-
d that there were three key features
sure nursing home experiences:

, report (occurrence or frequency of
rception of event)
example, single day, multiple days,

); frequency reports (e.g., "always"
quency (e.g., "every day, some days,

ed adjectives (e.g., "excellent" to
ion, numbered rating.

pts (such as food, getting help, and
varied all the question characteristics
possible options, the team was able
these questions. Appendix B shows
ns that could be asked about one con-
sed in the next rounds of cognitive
ds two and three was a systematic test
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 question, 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 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, the team refined the testing to focus on the various types of response tasks. The team tested rating with 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 card. 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 the 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 from 0 to 10 than 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.

Rounds 4 and 5. Once the question format was decided, these rounds focused on question content and wording. In contrast 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 from 0 to 5). These rounds also tested the vignettes (described later in this article). Testing rounds 6 and 7 are described as follows.

Material may be protected by copyright law (Title 17, U.S. Code)

Merging QOC and 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 (2004) 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 including questions used in the hospital CAHPS instrument, the in-center hemodialysis CAHPS instrument, and, of course, the fall 2003 NHCAHPS instrument (RTI, 2003). This analysis revealed that many 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 (Schipper, Clinch, & Olweny, 1996). 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, spirituality) 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 QOC 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 used 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 an 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 selecte
learned from cognitive testing th

- *Response tasks.* The 0 to 1
possible) did not work for n
no" also was not an adequat
No/Sometimes" was tested
QOL items.
- *Order/structure.* Respondei
switch back and forth betwe
yes/Mostly no" questions, a
The order of the questions w
questions first. This worked
7 of testing.
- *Screeners.* Some items that
swer, such as being left lyin
actually needed screening qu
turn/move oneself in bed).

Additional edits were made, ba
the QOL items. In May 2005, a pr
completed. The pretest provided i
bined instrument worked togethe
tion about the actual protocol use

*Development and Testing of Vig
as a Potential Cognitive Screen*

As previously discussed, throug
cess, the team was very concerne
could not participate in the NHC
screeners were used, yet none test
answer the questionnaire.

The team reviewed the literatu
and long-term memory, ability to
and recall. They examined and con
(Folstein, Folstein, & McHugh, 1
Gray, 1992), Short Blessed (Katzma
clock drawing tests (Ferrucci et a
Neill & Lichtenberg, 2000), the C

Material may be protected by copyright law (Title 17, U.S. Code)

cts

ıe nursing home resident experience
QOL combined, the team had to se-
items developed by Kane and col-
ı with the NHCAHPS QOC items.
in and item content of the QOL mea-
measures that were currently under
ns that are frail and require intense
he hospital CAHPS instrument, the
rument, and, of course, the fall 2003
3)). This analysis revealed that many
sures actually tapped QOC and were
n other CAHPS instruments. That is,
instruments referred to aspects of
ıe actual items included in some of
g., whether care providers communi-
The concept of "QOL" is very broad
sciplines and perspectives. But what
he idea that QOL is a subjective state
eny, 1996). On the other hand, QOC
he care delivered. The two concepts
ıat QOC refers to health care process
QOL refers to health care outcomes
) whom care was delivered as an out-

ed the content of the QOL items to
) QOL (e.g., autonomy, spirituality)
communication with staff). The ulti-
ıat should be used to supplement the
ntify items that referred uniquely to
HPS survey. To help decide which
d several criteria, including whether
g home quality improvement, what
ke, what the item's relationship was
l QOL rating, and whether the item
rsing homes.
ıterviewing, we focused on the QOL
ınderstanding of the new items and
n, we tested if there are any order

effects of QOL items and selected QOC items. Some of the key findings learned from cognitive testing the QOL items were:

- *Response tasks.* The 0 to 10 rating scale (worst 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 was 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 put 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 one 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.

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 (Folstein, Folstein, & McHugh, 1975), EXIT25 (Royall, Mahurin, & Gray, 1992), Short Blessed (Katzman et al., 1983), CLOX 1 and additional clock drawing tests (Ferrucci et al., 1996), animal naming test (MacNeill & Lichtenberg, 2000), the Cognitive Performance Scale or CPS

Material may be protected by copyright law (Title 17, U.S. Code)

(Morris et al., 1994), 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 (Schmand et al., 1999). 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) on rating of help with dressing with the same 0 to 10 rating task used for QOC items. These vignettes were administered by interviewers as part of the survey and were never used as screeners (i.e., interviews 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, give 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 (Katzman et al., 1983). 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
score generally provided additional
those better able to answer from those
within a CPS score of 1, persons with
in-range responses on average, while
only 81% in-range responses on avera
error scores were more consistent wi
CPS scores. Residents judged by inter
probes at all had higher mean vignette
rors in scaling the vignettes) compared
ers as understanding probes "sometim
for both groups and did not differenti

In summary, the vignettes used in t
to enhance the ability of the CPS to id
higher percentages of questions. The
relationship with interviewer confide

Field Test

The field test was conducted from
nursing homes in the states of Maine
cut. The 12 nursing homes were recru
ing homes in these three states and by
a mix of non-profit and proprietary,
and small/medium (40 to 114 beds) an

Appendix D shows the domains rep
instrument used in the field test and t
instrument versions were tested: An in
ministered to current residents who h
did not have a planned discharge (n
short-stay (90 days or less) residents d
123). Each nursing home provided a
discharged residents. Eligibility crite
staff, such as excluding residents with
to be in a coma, or "severely impaired
then randomly drawn from the list o
mail survey was sent primarily to pos
had a question about therapy provide
about eye, dental, and hearing aid ca

Material may be protected by copyright law (Title 17, U.S. Code)

letermine which might best be suited
ot able to find a short screener that
Many of the short standardized as-
itation; yet, orientation to place and
nursing home residents. Moreover,
information on how predictive tem-
ther orientations.
gnitive screening identified through
method (Schmand et al., 1999). The
of residents' abilities to generalize
ences and to assign a numeric rating
team felt that the vignettes should be
esidents might need help. We devel-
ndix C) on rating of help with dress-
; task used for QOC items. These
erviewers as part of the survey and
nterviews were never terminated be-
ered). Instead, the vignette responses
ere completed and compared to re-
es and to the survey answers. Resi-
ies to score the vignettes in a logical
a vignette in which the depicted per-
ng than to one with help most of the
nette with help most of the time than
The responses to the three vignettes
sponse pattern was as described.
, the interviewers administered the
overs short-term memory, temporal
n et al., 1983). In addition, the inter-
neir perceptions of residents' under-
the questions and of the cognitive
ed for each resident in order to com-
r measures.
lictive of the percentage of questions
at, compared to others, respondents
lso tended to have lower percentages
example, respondents with three er-
77% of questions in the survey, on
nts with no errors on the vignettes
the survey. The team also looked at
wered 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 zero vignette errors gave 100% in-range responses on 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 than 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.

Field Test

The field test was conducted from June through August 2005 in 12 nursing homes in the states of Maine, New Hampshire, and Connecticut. The 12 nursing homes were recruited from letters sent to 206 nursing homes in these three states and by personal contact. The homes were a mix of non-profit and proprietary, freestanding and hospital-based, and small/medium (40 to114 beds) and large (115 or more beds).

Appendix D shows the domains represented in the resident NHCAHPS instrument used in the field test and the specific questions asked. Two instrument versions were tested: An in-person interview instrument administered to current residents who had a stay of 30 or more days and did not have a planned discharge (n = 424), and a mail version for short-stay (90 days or less) residents discharged in the past 60 days (n = 123). Each nursing home provided a census of its current and recently discharged residents. Eligibility criteria were applied by the research staff, such as excluding residents with MDS records that showed them to be in a coma, or "severely impaired in decision making." Sample was then randomly drawn from the list of eligible residents. Because the mail survey was sent primarily to post-acute care patients, this version had a question about therapy provided and did not include questions about eye, dental, and hearing aid care. The resident instrument and

Material may be protected by copyright law (Title 17, U.S. Code)

survey protocols have been revised based on the psychometric and implementation analyses of the field test in 2006.

LESSONS LEARNED

The resident NHCAHPS development 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 (Harris-Kojetin et al., 1999). Because of repeated evidence that residents had trouble with reference periods, our recommendation is to use the non-specific present, in contrast with typical survey methodology and other CAHPS surveys where explicit time reference periods are used (Harris-Kojetin et al., 1999). The NHCAHPS testing found that the 0 to 10 response scale appeared to work well with nursing home residents for many of the QOC questions. This use of the 0 to 10 scale is consistent with other CAHPS surveys (Harris-Kojetin et al., 1999) and some other research with the elderly (Castle & Engberg, 2004). Our testing did find, however, that a different response scale ("yes/sometimes/no") was needed for many of the QOL items.

Creating a taxonomy of possible question formats and testing the formats before the content may be a model for future CAHPS® surveys that focus on consumers or patients with limited cognitive abilities. NHCAHPS is the first CAHPS survey to include QOL items in addition to the usual QOC items. The QOL testing results might provide guidance for survey development in other settings that combine both QOC and QOL, such as assisted living.

The team's work on the vignette method holds promise as a method to select residents for surveys who can provide reliable responses. In the field test, about one-third of residents who successfully completed in-person interviews had CPS scores of 3 or more, a group often excluded from survey participation. In that respect, this percentage is similar to that found in another study using a three-approach interviewer protocol (van Maris et al., 1996).

Material may be protected by copyright law (Title 17, U.S. Code)

based on the psychometric and im-
est in 2006.

LEARNED

ɔment demonstrates the critical role
vey items with the intended respon-
ation, particularly for a population
nursing home residents. The cogni-
n understand the most appropriate
guidance on types of questions, time
esponse task. In contrast to other
m concluded that ratings were more
ffieulty that residents had with sum-
ris-Kojetin et al., 1999). Because of
trouble with reference periods, our
ɔecific present, in contrast with typi-
CAHPS surveys where explicit time
ojetin et al., 1999). The NHCAHPS
se scale appeared to work well with
he QOC questions. This use of the 0
AHPS surveys (Harris-Kojetin et al.,
ɪe elderly (Castle & Engberg, 2004).
different response scale ("yes/some-
he QOL items.
e question formats and testing the
model for future CAHPS® surveys
ts with limited cognitive abilities.
ey to include QOL items in addition
testing results might provide guid-
er settings that combine both QOC

method holds promise as a method
ın provide reliable responses. In the
ents who successfully completed
es of 3 or more, a group often ex-
that respect, this percentage is sim-
ısing a three-approach interviewer

AHRQ has established a listserve as a vehicle for sharing information about the development of these instruments and future instruments relating to other long-term care settings (e.g., assisted living), as well as exchange of information relating to research and related events. Questions can be sent to LTCCAHPS@ahrq.gov for more details.

RECEIVED: 01/05
REVISED: 01/05
ACCEPTED: 01/06

REFERENCES

Allen, I., Hogg, D., & Peace, S. (1992). *Elderly People: Choice, Participation and Satisfaction*. London: Policy Studies Institute.

Castle, N. & Engberg, J. (2004). Response formats and satisfaction surveys for elders. *The Gerontologist, 44*(3), 358-367.

Edwards, W., Silva, P., & Narayanan, V. (2000). *Resident Experience with Nursing Home Care: A Literature Review*. Report Submitted to AHRQ/DHHS.

Ferrucci, L., Cecchi, F., Guralnik, J. M., Giampaoli, S., Lo Noce, C., Salani, B., Bandinelli, S., & Baroni, A. (1996). Does the clock drawing test predict cognitive decline in older persons independent of the mini-mental state examination? *Journal of the American Geriatrics Society, 44*, 1326-1331.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Fowler, F. J. (1995). *Improving Survey Questions*. Thousand Oaks, CA: Sage Publications, Inc.

Harris-Kojetin, L., Fowler, F., Brown, J., Schnaier, J., & Sweeny, S. (1999). The use of cognitive testing to develop and evaluate CAHPS 1.0 core survey items. *Medical Care, 37*(3), MS10-MS21.

Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.

Kane, R. A., Kane, R. L., Giles, K., Lawton, M. P., Bershadsky, B., Degenholtz, H., Kling, K., Cutler, L., & Grant, L. (2000). *First Findings from Wave 1 Data Collection: Measures, Indicators, and Improvement of Quality of Life in Nursing Homes*. Report to CMS, Prepared under Master Contract #500-96-0008.

Kane, R. A., Kling, K., Bershadsky, B., & Kane, R. L. (2003). Quality of life measures for nursing home residents. *Journal of Gerontology: Medical Sciences, 58A*, M240-M248.

Kane, R. A., Kane, R. L., Bershadsky, B., Cutler, L., Giles, K., Kling, K., & Degenholtz, H. (2004). *Measures, Indicators, and Improvement of Quality of Life in Nursing Homes*. Final Report to CMS, Prepared under Master Contract #500-96-0008.

Material may be protected by copyright law (Title 17, U.S. Code)

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal of Psychiatry, 140*, 734-739.

MacNeill, S. E., & Lichtenberg, P. A. (2000). The MacNeill-Lichtenberg decision tree: A unique method of triaging mental health problems in older medical rehabilitation patients. *Archives of Physical and Medical Rehabilitation, 81*, 618-622.

Morris, J. N., Fries, B. E., Meher, D. R., Hawes, C., Phillips, C. D., Mor, V., & Lipsitz, L. A. (1994). MDS cognitive performance scale. *Journals of Gerontology, 49*, M174-M182.

Mozley, C., Huxley, P., Sutcliffe, C., Bugley, H., Burns, A., Challis, D., & Cordingley, L. (1999). "Not knowing where I am doesn't mean I don't know what I like": Cognitive impairment and quality of life responses in elderly people. *International Journal of Geriatric Psychiatry, 14*, 776-783.

Myers, F. & MacDonald, C. (1996) "I was given options not choices": Involving older users and carers in assessment and care planning. In R. Bland (Ed.), *Developing Services for Older People and Their Families*. London: Jessica Kingsley.

Phillips, C. D., Chu, C. W., Morris, J. N., & Hawes, C. (1993). Effects of cognitive impairment on the reliability of geriatric assessments in nursing homes. *Journal of the American Geriatrics Society, 41*(2), 136-142.

Royall, D. R., Mahurin, R. K., & Gray, K. F. (1992). Bedside assessment of executive cognitive impairment: The executive interview. *Journal of the American Geriatrics Society, 40*, 1221-1226.

RTI International, Harvard University, RAND Corporation (2003). *CAHPS Instrument for Persons Residing in Nursing Homes*. Final Report to CMS, CMS Contract No. #CMS-01-01176.

Sansone, P., Schmitt, L., Nichols, J., Phillips, M., & Belisle, S. (1998). Determining the capacity of demented nursing home residents to name a health care proxy. *Clinical Gerontologist, 9*(4), 35-50.

Schipper, H., Clinch, J. J., & Olweny, C. L. M. (1996). Quality of life studies: Definitions and conceptual issues. In B. Spilker (Ed.) *Quality of Life and Pharmacoeconomics in Clinical Trials*. Second Edition. Philadelphia, PA: Lippincott-Raven Publishers. pp. 11-23.

Schmand, B., Gouwenberg, B., Smit, J., & Jonker, C. (1999). Assessment of mental competency in community-dwelling elderly. *Alzheimer Disease and Associated Disorders, 13*(2), 80-87.

Simmons, S., Schnelle, J., Uman, G., Kulvicki, A., Lee, K., & Ouslander, J. (1997). Selecting nursing home residents for satisfaction surveys. *The Gerontologist, 37*(4), 543-550.

van Maris, B., Soberman, L., Murray, M., & Norton, P. (1996). Satisfaction of residents and families in long-term care: II. Lessons learned. *Quality Management in Health Care, 4*(3), 47-53.

Willis, Gordon B. (2004). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications, Inc.

doi:10.1300/J031v19n02_04

APPEN
Steps in Resident NH

Phase 1: Resident Survey: Feasibility

1. Literature Review completed May 2000
2. Interviews with experts*
3. Methodological Expert Group (MEG) m

Phase 2: Resident Survey: Instrument De

1. Focus groups with residents and family
2. Cognitive testing for Quality of Care OC
   a. Round 1: Fall 2001 (52 residents in
   b. Round 2: Spring 2002 (15 residents
   c. Round 3: Spring 2002 (19 residents
   d. Round 4: Spring 2003 plus vignette
   e. Round 5: Summer 2003 plus vignett
3. Draft Instrument for Resident NHCAHP
4. CMS Decision to Merge Kane's Quality
5. Selection of QOL items: Fall 2003-Sum
6. Cognitive testing for Quality of Life item
   a. Round 6: Fall 2004-Winter 2005 (16
   b. Round 7: Spring 2005 (19 residents

Phase 3: Resident Survey: Field Test

1. Pretest of Merged Instrument (May 200
2. Field Test (June-August 2005)

Note: *Steve Albert, Kitty Buckwalter, Tim Case Johnson, Rosalie Kane, Powell Lawton, Vince Mo Phil Sloan, Joan Teno, Gwen Uman, Sheryl Zimm

**Robert and Rosalie Kane; Farida Ejaz; Catherine Powell Lawton; Jay Magaziner; Vincent Mor; Rud Soberman; Joan Teno; Sheryl Zimmerman.

Material may be protected by copyright law (Title 17, U.S. Code)

APPENDIX A
Steps in Resident NHCAHPS Development

Phase 1: Resident Survey: Feasibility

1. Literature Review completed May 2000
2. Interviews with experts*
3. Methodological Expert Group (MEG) meeting** (June 2000)

Phase 2: Resident Survey: Instrument Development

1. Focus groups with residents and family members: Spring 2001
2. Cognitive testing for Quality of Care QC items
   a. Round 1: Fall 2001 (52 residents in five homes)
   b. Round 2: Spring 2002 (15 residents in three homes)
   c. Round 3: Spring 2002 (19 residents in three homes)
   d. Round 4: Spring 2003 plus vignette testing (27 residents in three homes)
   e. Round 5: Summer 2003 plus vignette testing (31 residents in three homes)
3. Draft Instrument for Resident NHCAHPS: Fall 2003
4. CMS Decision to Merge Kane's Quality of Life (QOL) items from Study: Fall 2003
5. Selection of QOL items: Fall 2003-Summer 2004
6. Cognitive testing for Quality of Life items
   a. Round 6: Fall 2004-Winter 2005 (16 residents in two homes)
   b. Round 7: Spring 2005 (19 residents in two homes)

Phase 3: Resident Survey: Field Test

1. Pretest of Merged Instrument (May 2005)
2. Field Test (June-August 2005)

Note: *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.

**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; Sheryl Zimmerman.

Material may be protected by copyright law (Title 17, U.S. Code)

APPENDIX B
Possible Variations for "Staff Comes Quickly When Called" Concept

	Time Period		
	One Day (Yesterday/Today)	Week (Last 7 days/Last week)	Non-Specified Present
Reports			
Yes/No	Yesterday, did someone come quickly when you called for help?	In the last week, did someone come quickly when you called for help?	Does someone come quickly when you call for help?
Frequency report with adverbs such as: Always to Never	Yesterday, how often did someone come quickly when you called for help?	In the last 7 days, how often did someone come quickly when you called for help?	How often does someone come quickly when you call for help?
Days-based frequency task such as: 1) How many days? or 2) Every days, most days, some days, no days	—NOT POSSIBLE—	In the last 7 days, on how many days did someone come quickly when you called for help?	How often does someone come quickly when you call for help?
Ratings			
Ratings with adjectives such as: Excellent to Poor	Yesterday, how would you rate how quickly someone came when you called for help?	In the last 7 days, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?
Comparative evaluation such as: 1) needs no improvement, needs a little improvement, needs a lot of improvement or 2) OK, could be a little better, could be a lot better	Yesterday, how would you rate how quickly someone came when you called for help?	In the last week, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?
Numbered ratings such as: 0 to 10 (use any number from 0 to 10 where 0 is the worst possible and 10 is the best possible.)	What number would you use to rate how quickly someone came when you called for help today?	Thinking about the last week, what number would you use to rate how quickly someone came when you called for help?	What number would you use to rate how quickly someone comes when you call for help?

APPEN
Vignettes Used in Cognitive

VIGNETTES (Cognit

Interviewer Preface: I'm going to tell you abo
and ask you to give a score to the help they g
the worst.

1. Mr. Baker is in a nursing home. He likes to
dressed by himself and needs help. Staff mem
him get dressed so he stays in his pajamas.

Using any number from 0 to 10 where 0 mean
possible help, what number would you give th

2. Let me ask you about a different person. Mr
day and she needs help. **Most days,** when th
get dressed the way she likes. Once a week w
her get dressed at all and she stays in her nig
She is usually dressed, but sometimes not.

Using any number from 0 to 10 where 0 mean
possible help, what number would you give to

3. Mr. Jones is the third person. He also likes
help. **Every day,** staff members come to his r
likes.

Using any number from 0 to 10 where 0 mean
possible help, what number would you give th

VIGNETTES (Fie

[HAND R SHOW CARD #1: 0-10] Interviewe
you can answer using any number from 0 to 1
the best possible. (POINT TO NUMBERS).

Let's start with three practice questions. I'm go
who live in a nursing home and ask you to rate
All three people want to get dressed every day

[IF NEEDED: These people are only examples
made-up and are not real.]

1. The first person is Mr. Baker. Staff is always
dressed so he stays in his pajamas.

Using any number from 0 to 10 where 0 means
ble, what number would you use to rate the he

2. The second person is Mrs. Smith. Every day

What number would you use to rate the help M

3. Mr. Jones is the third person. **Most days** sta
when staff are too busy, they do not help him g
one day.

What number would you use to rate the help M

Material may be protected by copyright law (Title 17, U.S. Code)

or 2) Every days, most days, some days, no days			when you called for help?
Ratings			
Ratings with adjectives such as: Excellent to Poor	Yesterday, how would you rate how quickly someone came when you called for help?	In the last 7 days, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?
Comparative evaluation such as: 1) needs no improvement, needs a little improvement, needs a lot of improvement or 2) OK, could be a little better, could be a lot better	Yesterday, how would you rate how quickly someone came when you called for help?	In the last week, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?
Numbered ratings such as: 0 to 10 (use any number from 0 to 10 where 0 is the worst possible and 10 is the best possible.)	What number would you use to rate how quickly someone came when you called for help today?	Thinking about the last week, what number would you use to rate how quickly someone came when you called for help?	What number would you use to rate how quickly someone comes when you call for help?

APPENDIX C
Vignettes Used in Cognitive Interviewing and Field Test

VIGNETTES (Cognitive Testing Version)

Interviewer Preface: I'm going to tell you about three different people in a nursing home and ask you to give a score to the help they get from staff. 10 is the best score and 0 is the worst.

1. Mr. Baker is in a nursing home. He likes to get dressed every day but he can't get dressed by himself and needs help. Staff members are **always too busy** and **never help him** get dressed so he stays in his pajamas.

Using any number from 0 to 10 where 0 means worst possible help, and 10 means best possible help, what number would you give the help Mr. Baker gets getting dressed?

2. Let me ask you about a different person. Mrs. Smith also likes to get dressed every day and she needs help. **Most days,** when the aides aren't busy, they come to help her get dressed the way she likes. Once a week when the aides are too busy, they don't help her get dressed at all and she stays in her nightgown all day. **(Once a week, if asked)** She is usually dressed, but sometimes not.

Using any number from 0 to 10 where 0 means worst possible help, and 10 means best possible help, what number would you give to the help Mrs. Smith gets getting dressed?

3. Mr. Jones is the third person. He also likes to get dressed each day and he needs help. **Every day,** staff members come to his room and help him get dressed the way he likes.

Using any number from 0 to 10 where 0 means worst possible help, and 10 means best possible help, what number would you give the help Mr. Jones gets getting dressed?

VIGNETTES (Field Test Version)

[HAND R SHOW CARD #1: 0-10] Interviewer Preface: For the first group of questions, you can answer using any number from 0 to 10 where 0 is the worst possible and 10 is the best possible. {POINT TO NUMBERS}.

Let's start with three practice questions. I'm going to tell you about three different people who live in a nursing home and ask you to rate the help they get from staff with dressing. All three people want to get dressed every day but cannot get dressed by themselves.

[IF NEEDED: These people are only examples. They're not in this nursing home. They're made-up and are not real.]

1. The first person is Mr. Baker. Staff is always too busy and **never help him get dressed** so he stays in his pajamas.

Using any number from 0 to 10 where 0 means worst possible, and 10 means best possible, what number would you use to rate the help Mr. Baker gets with his dressing?

2. The second person is Mrs. Smith. Every day staff members help her get dressed.

What number would you use to rate the help Mrs. Smith gets with her dressing?

3. Mr. Jones is the third person. **Most days** staff help him get dressed. But once a week, when staff are too busy, they do not help him get dressed so he stays in his pajamas that one day.

What number would you use to rate the help Mr. Jones gets with his dressing?

Material may be protected by copyright law (Title 17, U.S. Code)

APPENDIX D
NHCAHPS Questions in Field Test Version (Summer 2005)

Domains	Questions (text shortened for space)
Rating of NH experience/ environment	(1) Food*; (2) mealtimes*; (3) temperature*; (4) cleanliness; (5) noise levels (day and night); (6) how room is set up#
Getting needed care from NH	(1) How well medicine worked to help with pain; (2) how well staff help you when you have pain #; (3) how gentle staff ; (4) left sitting or laying in the same position so long that it hurts #; (5) (mail survey only) special therapy received
Getting care without long wait	How quickly staff come when you call for help #
Courteous, respectful treatment	How respectful staff are #
Communication with nurses and nurses aides	(1) How well staff listen; (2) how clearly staff explain things
Getting needed care from doctors	Rating of care from any doctors
Global Care Ratings	Overall care from staff
Overall NH rating	(1) Rating of nursing home; (2) recommend this nursing home to others?
Eye, Ear, and Dental Care (current resident survey only)	(1) How well staff keep hearing aid in good working condition; (2) how well staff help with eye care; (3) how well staff help with dental care
QOL (Security; Religious Needs; Privacy; Autonomy/ Choice; Meaningful Activity)	(1) How safe and secure*; (2) how well NH helps meet religious needs*; (3) able to ability to find private place to visit*; (4) staff ensure personal privacy*; (5) ability to choose time to go to bed*; (6) ability to choose clothes*; (7) ability to choose activities*; (8) volunteers or staff talk about activities*; (9) have enough organized activities (weekdays and weekends)*
Global QOL rating	Rate your life now*

*Item is one of or derived from Kane QOL items
#Item overlaps CAHPS concept and Kane QOL items
Unmarked items are considered CAHPS concepts

Developing and Tes
Survey for Nursing
The Ohio E

Jane K. Str
Miami Universi

Farida K. E
Catherine M
Margaret Blenkner Research Institute of the

James A. Jo
Ball State Univers

ABSTRACT. Input from consum
quality improvement in long-term ca
This paper documents the developm
dent Satisfaction Survey (ONHR

Jane K. Straker, PhD, is Director of Polic
ect, Scripps Gerontology Center, Miami Un
45056 (E-mail: strakejk@muohio.edu).
Farida K. Ejaz, PhD, is Senior Research
Institute of the Benjamin Rose Institute, 119
OH 44120 (E-mail: fejaz@benrose.org).
Catherine McCarthy is Senior Research
Institute of the Benjamin Rose Institute (E-m
James A. Jones, PhD, is Assistant Directo
puting Services, Ball State University, Munci
Address correspondence to: Jane K. Strake
Term Care Research Project, Scripps Geron
Upham Hall, Oxford, OH 45056 (E-mail: stra

Journal of Aging & Social P
Available online at http://ja
© 2007 by The Haworth Press
doi:10.1300/J031

Material may be protected by copyright law (Title 17, U.S. Code)

Nursing Home CAHPS Field Test Report

Carol Cosenza
Floyd J. Fowler, Jr.
Center for Survey Research
University of Massachusetts Boston

Joan L. Buchanan
Paul D. Cleary
Lin Ding
Harvard Medical School

1-19-06

The work described in this report was sponsored through an Intra-agency agreement between AHRQ and the Centers for Medicare & Medicaid Services (CMS). The project was funded by a cooperative agreement between the Agency for Healthcare Research and Quality and the Harvard Medical School (Grant # U18HS09205). CMS was a full partner in the project design and execution.

Nursing Home CAHPS Field Test Report
Executive Summary

I. BACKGROUND

The Centers for Medicare & Medicaid Services (CMS) is interested in using data from surveys of home residents and their families to assess nursing home quality. In 1995, the Agency for Healthcare Research and Quality (AHRQ) started a research program to develop patient-based measures of care experiences, referred to as the Consumer Assessment of Healthcare Providers and Systems (CAHPS) project. CMS and AHRQ have worked with members of the CAHPS consortium to develop and test surveys that can be used to collect data from nursing home residents about their care experiences.

This report 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.

The pilot study was carried out by the Center for Survey Research at the University of Massachusetts Boston (CSR) under subcontract with Harvard Medical School. The CMS Privacy Board approved a waiver of individual authorization for this NHCAHPS research project on April 22, 2005.

II. IN-PERSON INTERVIEWS OF LONG-TERM RESIDENTS

A. PURPOSES

There were several goals of the field test including:

1. Evaluating protocols for developing probability samples of nursing home residents
2. Estimating the rates at which nursing home residents would be eligible, and be able and willing to be interviewed
3. Evaluating a data collection protocol, including staffing, managing the data collection, and identifying residents who are able to provide meaningful answers to questions
4. Estimating the time and effort required to field such a project.

B. DESIGN AND PROCEDURES

Letters were sent to 206 diverse nursing homes in Connecticut, Maine, Massachusetts, and New Hampshire. Follow-up telephone calls were made to a subset of these. When the sample composition target was met, 19 nursing homes had agreed to cooperate, and, of them, 13 eventually participated. One home was used for pretesting procedures and the other 12 were used in the main field test.

SAMPLE:

Eligible respondents were residents who had been in the home for at least 30 days with no discharge planned. We excluded residents whose MDS assessment indicated that they were "severely impaired in cognitive skills for daily decision making", those who were comatose, and those for whom a legal guardian was required to make medical decisions.

PROTOCOL:

The target number of interviews was 25 for homes with fewer than 115 beds and 40 for homes with 115 or more beds. The protocol was to randomize a list of eligible residents, and then proceed down the list in order until the target number of interviews was completed. Data collection took place over 2 days in each nursing home.

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. Interviews were conducted between June 22, 2005 and August 9, 2005.

PRETEST

On May 26, 2005, the protocol and survey were pretested. Based on our sampling criteria, over half of the pretest sample was found to be ineligible. We also found 3 additional cases that were ineligible at the nursing home. Overall, we had an eligibility rate of 53.7%. Of the 31 eligible residents, we were able to interview 15 (48.4%).

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

C. FIELD TEST RESULTS

1. 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.

2. 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.

3. 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.

4. 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.

5. 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

D. TIME AND EFFORT REQUIRED FOR CURRENT RESIDENT PROTOCOL

The estimates below do not include any travel time or costs, (nor recruiting time), since those will be dependent on the location of the project staff, interviewers, and sampled nursing homes. They also do not include any information about the time required to recruit the nursing homes.

1. Sampling

Sample collection took about 137 hours of effort. Most of this - 72 hours - was done by interviewers and clerical level staff. The rest of the time was spent by the project manager and research assistant gathering, organizing, and editing the sample data from the nursing homes.

The cost for a larger data collection effort would depend on what percentage of homes could deliver electronic data files. If our experience is representative, we estimate that identifying a sample would average 6 hours per home in clerical staff time and about half a day per home of professional staff time.

2. Data Collection

It took approximately 1178 hours of staff time to field this study. This number does not include the costs and time associated with travel, nor does it include time spent on basic (non-study specific) interviewer training. This averages to about 3 hours of staff time per completed interview. For this field test we used professional interviewers. This allowed us to be very efficient in the use of interviewing time and gave us confidence that the interview was being administered in a consistent and standardized way. Having interviewers who understood the interview process also helped make sample decisions, such as when a respondent was not able to continue the interview, a less arbitrary event. Using non-professional interviewers would change not only the cost implications, but most likely the consistency of interviewer decisions and the quality of the data they collected.

3. Summary - Cost Estimates

If one excludes travel costs, thereby including only the time when staff members were in the homes, the direct costs for data collection were about $24,000 or about $57 per interview.

In addition to travel, other issues to consider when extrapolating from this experience to possible future surveys:

1. Pay rates could differ.
2. On a per interview basis, the effect of some of the starting costs might be lower for an ongoing data collection.
3. Because of the level of control we wanted to exercise, we probably had more supervision than would be necessary in an ongoing survey.
4. If the interview is shortened by dropping items or one or both of the cognitive evaluation series, costs per interview might drop a little.
5. These figures do not include indirect costs, which will be specific to the data collection organization.

We also calculated costs per home and per interview for large and small homes. There were costs, such as interviewer training, sampling and setting up the home, that were unaffected by home size. These fixed costs came out to about $790 per home for 12 homes. The data collection costs per home were basically proportionate to the number of interviews: At about $30 per interview for interviewing and supervision, small homes cost about $846 for an average of 28 interviews and large homes cost about $1150 per home for an average of 41 interviews. Sample control and data entry were also proportionate to the number of interviews. The costs were about $5.60 per interview. This came out to about $155 for each small home and $230 for each large one. All of these estimates exclude travel and overhead.

E. FEEDBACK ABOUT CURRENT RESIDENT PROTOCOL

1. 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. Administrators had high praise for the interviewing team, their friendliness, and their self-sufficiency. On the whole, there were no disruptions or difficulties.

2. Feedback from Interviewers

Overall, the interviewers thought the process went very well and enjoyed doing the interviews. They all felt that having a "team" with an on-site coordinator was the most efficient protocol. They felt without that, it would have been more of a burden on the nursing home staff. Having the show cards was useful. Amplifiers (for respondents who had difficulty hearing) were not used a lot but interviewers said that having them was very helpful.

The interviewers said that the Short Blessed was often awkward to administer and was hard to score. Furthermore, it increased the length of the interview. Many interviewers said that it was a bad way to end the interaction. Administering the vignettes was also difficult. Interviewers thought that some respondents were confused by the hypothetical situations.

3. 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.

II. MAIL SURVEY OF SHORT-TERM DISCHARGED RESIDENTS

A. BACKGROUND & PURPOSE

Short-term nursing home residents' stays are usually primarily a medical event, not a total living experience. The cognitive problems that are so prevalent among long-term residents are much less prevalent in the short-term population.

The survey instrument for long-term residents was adapted slightly (adding a question about therapy received and deleting questions about help with hearing aids, and eye and dental care) and put into a mail survey form for the short-term residents. For convenience and timing, the short-term residents of the nursing homes participating in the long-term field test were used as the sample frame for this effort.

The main motivation for this phase of the pilot study was to learn about eligibility and collect some data from this population that could be compared with the results from the survey of long-term residents.

C. DESIGN & PROCEDURES

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

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

PROTOCOL:
The protocol was to send an initial mailing, with a cover letter, fact sheet, copy of a self-administered questionnaire, and a postage-paid return envelope (Appendices I and J). A packet with a second questionnaire was to be sent after two weeks to the non-responders (Appendix K). Then, nonresponders were to get a reminder call to make sure they had received the questionnaire, answer any questions, and urge participation.

The first mailing was on June 28. Unfortunately, 5 nursing homes could not provide their completed samples until well into July. Because of a hard deadline for data collection of mid-August, there was not time for follow-up mailings to those nursing homes. Those homes received their first (and only) mailing at the same time as the other 6 received their second mailing. One nursing home was not recruited until the end of July. Although in-person interviews were collected, this home did not participate in the mail portion of the field test.

CSR called everyone who had not yet returned a completed questionnaire to prompt people to return the written survey. Telephone reminder calls were conducted by professional interviewers. Each case had up to 3 calls made on different days and at different times of the day to attempt to make contact with the appropriate respondent.

D. FINDINGS

1. Sampling and Eligibility

The initial sample consisted of 381 residents from 11 nursing homes. 133 residents were ineligible because they did not meet the eligibility criteria. The major reason for ineligibility was being discharged to another care facility or being deceased. Very few in the sample were ineligible because of cognitive impairment or having a legal guardian/legal oversight. In order to obtain sufficient levels of response for reporting results by individual nursing home, future research might consider different sampling options, including using a rolling sample, where residents over several months are surveyed, rather than just the 2 month window that was used for the field test.

2. Data Collection

We found that the quality of the contact information was generally fairly good. About 70% of the sample sent from the nursing homes included phone numbers. Only one nursing home was unable to provide any phone numbers.

Overall, almost 52% of the eligible sample returned a completed survey. With time to implement a good mail protocol (including an option of telephone interview), this field test experience would suggest acceptable rates of return could be achieved. Discharged residents who received the more standard 2-mailings had a 57% response rate, while those that only received one mailing had a 43% response rate. There were very few explicit refusals to participate.

C. CSR TIME AND EFFORT (Discharged Resident Protocol)

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

III. ANALYSIS

Analyses of completeness and integrity of data

Socio-demographic information on residents collected from the Nursing Home Minimum Data Set (MDS) data was compared to similar items included in the survey for all respondents. Age discrepancies increased with cognitive impairment as measured by both the Cognitive Performance Scale (CPS) from the nursing home Minimum Data set and the Short Blessed administered during the interview. Education discrepancies occurred but appeared to result largely from differences in the response categories between the two data sources. We conclude that having multiple sources of information about socio-demographic characteristics helped us to identify mismatched records and MDS data errors and that multiple sources are most useful when response categories can be matched directly.

Analysis of Responders and Non-Responders and Sample Comparability

For the interview sample, we compared the potentially eligible but not interviewed residents (many were unassigned to interviewers, others failed to complete the interview) to those interviewed and found that they were similar with respect to gender, race, education, and payer mix. However, those not interviewed were older (average age of 84 vs. 81years old) and more cognitively impaired. There were no significant differences between responders and non-responders in the mail sample.

When we compared the interview sample and the mail responders, there were fewer African American mail respondents. The mail responders were significantly less cognitively impaired, and somewhat less likely to report feeling worried than the interview sample. There were small marginally significant differences in self-reported health status with mail respondents reporting somewhat worse health.

Item level applicability and response rates

We calculated the number of respondents who skipped each item because it was not applicable (skipped) and the number of respondents who did not answer an item that was applicable to them (no response). Hearing aid items and the item on being left sitting or laying too long in the same position had much lower applicability than other items; second highest were questions on dental and eye care – all items CMS requested.

Item non-response ranged from 1 to 6% for the interview sample. It was higher for the mail sample with most items ranging from 1 to 11%. The religious needs item had the highest non-response in both the interview (6%) and mail (18%) sample. Skipped pages contributed to these higher rates in the mail survey; 5 people skipped 1 page and 14 individuals skipped 2 pages.

Cognitive impairment and item responses

To examine the extent to which the different measures of cognitive functioning predicted completeness of response patterns, we first calculated an index of response completeness by calculating the number of inappropriately skipped items. We then examined the association between this index and the different measures of cognitive functioning. We found that 59% of interviewees answered all interview questions and 8% skipped four or more questions. As is usually the case, having an interviewer collecting data produces less item non-response than the mail survey. The fact that some people inadvertently skipped some pages also contributed to the difference. Only 25% of respondents answered every question in the mail survey and 27% skipped 4 or more items. The correlations between non-response and measures of cognition were statistically significant but weak, ranging from .10 to .29. For mail responders, the

correlation between the number of skipped items and cognition as measured by the CPS was not significant. Skipped pages in the mailed survey may account for this lack of relationship.

The Short Blessed scores had a stronger independent association with the number of unanswered questions than other measures of cognition (CPS, vignette score, and interviewer observation.) However, because the Short Blessed was administered at the end of the interview, we only have scores for individuals who were able to complete the interview whereas we have at least partial information on the other measures from everyone. We were not able to assess whether this stronger relationship with the Short Blessed would be achieved if we had been able to obtain scores Short Blessed scores on the entire sample.

Comparison of Response Patterns between Samples

The interview sample consisted of long stay nursing home residents and the mail survey was sent to individuals who had recently completed a short post acute care nursing home stay. In order to determine whether information should be reported in combination or separately for the two samples we compared the distribution of average scores across homes in the two samples. The overall distributions are similar in the two samples but the correlations between the two groups within each home are quite low for many of the items, including most of the 0-10 ratings, suggesting that the experiences of one group are not a good predictor the experiences of the other group. This could be because they have different needs and experiences, their expectations differ, and/or because of differences in their socio-demographic characteristics that might be related to reporting differences. Such differences should be evaluated in larger samples and when case-mix models have been developed. Those could be used to assess the extent to which the differences are due to differences between short and long stay patients. Until more data are available on this issue, caution should be used when combining information from these two groups of patients.

Ceiling Effects

The percent of respondents using the highest rating category for each item was reviewed. As expected, we found less clustering in the highest response category on the 0-10 scale than for the other response formats. For the interviewed sample using the 0-10 rating scale, in 7 out of 18 items more than 40 % of respondents used the highest rating. The same was true for only 3 of the 18 items in the mail respondent sample. Ceiling effects were more pronounced among items with only 3 response options (Yes, Sometimes, No). This was most pronounced in the autonomy items (choose bedtime, choose clothes, choose activities) and the personal privacy item.

Factor Analyses

We conducted exploratory factor analyses of all the questions (except for the global ratings) separately for mail and interview samples. Those analyses resulted in very discrepant patterns, with multiple factors in each sub-sample. Although the data suggest that the experiences of long and short stay residents are different and they potentially have different patterns of correlations, the sample of mail (short stay) residents was too small to develop precise estimates of factor analysis coefficients. Analyses of the combined data yielded a two factor solution. The factor pattern coefficients indicate that the responses clustered generally in ways that were anticipated, but there were no strong empirical grouping of items that were substantively consistent. Analyses of the interview only sample yielded similar results.

Given the lack of clear empirical clustering of items, we developed and tested several sets of scales, based primarily on the substantive content of the items. Because several items could be interpreted as reflecting more than one aspect of care, the study team reviewed different possible item combinations and considered the estimated reliability in the mail, interview, and combined samples as well as the face validity of each item for the scale considered. The scales selected as recommended reporting scales include a 9 item Environment scale, 5 item Care scale, 3 item Communication and Respect scale, 3 item Autonomy scale, and a 2 item Activities scale. The team also thought that the two food items included in the Environment scale could be treated as a separate subscale if there were interest in tracking that aspect of nursing home quality separately.

We calculated the correlations between the scales and several of the rating items. The most striking pattern is the large correlation between the Communication and Respect scale and the overall rating of care from nursing home staff. The scale that generally had the lowest correlations with the ratings is the Autonomy scale. The Environment and Communication and Respect scales tended to have the strongest correlations with the overall rating of the nursing home.

The composite scores tend to be highly correlated, indicating that they provide information that is to a large degree shared across scales. One possible inference is that residents general affect about their experience with the nursing home dominated their answers to individual questions. That is, if the only goal of the survey data were to develop scores that discriminated between nursing homes, without regard to the substantive content of items, it probably would be possible to develop a smaller number of composites that provided comparable information about inter-home variability.

The home level reliability of the scales for all subjects and for the separate samples indicate that with the sample sizes from the field test, most of the scales achieved only moderate reliability at the home level and that from 13 to 100 respondents would be needed for the scales to achieve a home level reliability of 0.70, depending on the sample, or mode of data collection. Since participation in the field test was voluntary, participating homes might be more homogeneous than a random sample of homes from more diverse locations.

If larger samples from a more diverse sample of homes confirm that more respondents are needed than many homes have at a particular time, it may be necessary to develop samples that are collected sequentially over longer periods of time. This strategy would allow nursing homes to accumulate adequate samples for making inferences about individual homes.

Conclusions and Recommendations

The Nursing Home CAHPS team successfully developed, refined, and tested procedures for conducting an interview survey of long term nursing home residents and a mail survey of short-stay discharged residents. Aside from the logistics of contacting the nursing homes, eliciting participation, drawing a sample, and arranging time to interview residents, the most challenging aspect of this project was collecting data from a group of individuals with a relatively high probability of cognitive impairment. Several instruments for testing the types of cognitive skills necessary to complete an interview were tested, and, the Short Blessed, was particularly successful at identifying residents who completed the interview but were not able to answer all of the questions. However, interviewers felt that administering the Short Blessed scale was burdensome and intrusive. The interviewers said that the strategy of stopping the interview whenever a respondent was unable to answer any three questions in a row was the least intrusive and stressful way to identify respondents who could not be interviewed. The team recommends that the CPS data in addition to this procedure be used in future interview studies. Other conclusions and recommendations regarding the implementation and costs of data collection activities are summarized in parts I and II of this report.

The instruments tested performed well. If they needed to be shortened, there were several questions that apply to a relatively small subset of respondents that could be dropped. In addition, there were several questions that the interviewers thought were interpreted differently than intended and/or that were difficult to answer, as described in Part I of this report. The team recommends that Question C30 (Ever left sitting or laying…) be dropped because it applied to so few residents and did not seem to fit well in any of the developed scales. Further work is needed to refine other questions, such as --C17 and C18 (getting care from doctor), C20 and C21 (help with hearing aid), C26 and C27 (religious support), C32 (how room was set-up), and C40 (asking about things to do). (These items and the issues that we became aware of during the data collection interviews are discussed in Section I of this report.) Of these items, only Q27 about religions needs is included in one of the recommended scales. However, the team thought that because of programmatic or constituent interest in the issues addressed by these questions, they should remain in the survey until alternative questions are developed and tested.

Nursing Home CAHPS Field Test Report

I. BACKGROUND

The Centers for Medicare & Medicaid Services (CMS) is interested in using data from surveys of nursing home residents and their families to assess nursing home quality. In 1995, the Agency for Healthcare Research and Quality (AHRQ) started a research initiative to develop patient-based measures of care experiences, referred to as the Consumer Assessment of Healthcare Providers and Systems (CAHPS) project. CMS and AHRQ have worked with members of the CAHPS consortium to develop and test surveys that can be used to collect data from nursing home residents about their care experiences. Once a draft survey was developed, CMS and AHRQ commissioned CAHPS consortium members to conduct a field test using that instrument.

The goals of the field test included learning 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 would differ for short and long term residents. Thus, the field test included two distinct activities. One focus was to test procedures for doing in-person interviews with long-term nursing home residents, those who had been in the home for 30 days or longer with no discharge planned. The second focus was to test a mail survey of recently discharged patients who had been in the home at least 5 days but fewer than 90 days, and who were discharged to their homes or to an assisted living facility.

The field test was carried out by the Center for Survey Research (CSR) at the University of Massachusetts Boston under subcontract with Harvard Medical School. Under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, covered entities are permitted to use or disclose protected health information for research purposes (45 CFR § 164.512(i)) provided that there is a Privacy Board approval of a waiver of individual authorization. The CMS Privacy Board approved a waiver of individual authorization for this NHCAHPS research project on April 22, 2005. The Privacy Board determined that Harvard Medical School and the University of Massachusetts-Boston met the criteria specified under HIPAA to obtain the waiver for NHCAHPS. CSR and Harvard also received Institutional Review Board/Human Subjects approval from their respective institutions for the field test.
Parts II and III of this report describes the pilot study and what was learned. Part IV presents analyses of the data collected.

II. IN-PERSON INTERVIEWS OF LONG-TERM RESIDENTS

A. PURPOSES

The specific goals of the field test included:

1. Evaluating protocols for developing probability samples of nursing home residents

2. Estimating the rates at which nursing home residents would be eligible to be interviewed, given a set of eligibility requirements

3. Estimating the rates at which eligible residents would be able and willing to be interviewed

4. Evaluating one approach to staffing and managing the data collection

5. Evaluating approaches to identifying residents who are able to provide meaningful answers to questions

6. Estimating the time and effort required to:
   a. Put together samples
   b. Collect data in homes
   c. Carry out other needed activities

7. Understanding how survey procedures worked from the perspective of nursing home staff

8. Gathering interviewer feedback on the survey instrument and individual questions

B. DESIGN AND PROCEDURES

For this field test, we wanted to include a variety of nursing homes – proprietary and not-for-profit, free-standing and hospital-based, as well as a mix of different sizes. Letters were sent to 206 nursing homes in Connecticut (n=124), Maine (n=34), Massachusetts (n=19), and New Hampshire (n=29). The locations of the homes were limited by a desire to use sites within about two hours' driving distance from Greater Boston, where CSR's interviewers lived. The project was on a tight time schedule, so it was necessary to find homes willing to give approval for participation quickly. Recruitment involved both project staff and staff from the sponsoring agencies. Follow-up calls were made to a subset of the 206 homes that received the original letter as well as several other homes where staff knew nursing home administrators personally or through other work relationships. When the sample composition target was met, 19 nursing homes had agreed to cooperate, and, of them, 13 eventually participated. One home was used for pretesting procedures and the other 12 were used in the main field test. Table 1 describes the 12 nursing homes in the field test.

Table 1: Nursing Home Characteristics

	For Profit	Not-for-Profit	Hospital-Based
Small/Medium (40-114 beds)	3	2	0
Large (115+ beds)	5	1	1

This sample is probably not representative of all nursing homes that would be willing to cooperate when there is more time to consider the request for participation. The homes that agreed to participate might have been among the better run and better organized homes. Our original time frame for this field test had assumed that all nursing homes would be recruited by May 20, 2005 (in order to have sufficient time for sample collection and coordination). Because of various complications, the final nursing home was not confirmed until July 22, 2005.

Once an administrator had agreed that a home would participate, a contact person at the home (often the administrator) was designated for all communication about the study. A CSR staff person contacted that person to confirm participation, help coordinate sample collection (or designate another person to do that), and to schedule interviewing dates. See Appendix A for the basic information that was collected about each nursing home prior to the visit.

We decided not to ask the nursing homes to sample patients, so that we could control how sampling was carried out and ensure that decisions about eligibility were made in a consistent way. Participating nursing homes were asked to provide 19 pieces of MDS information from the records of all the residents currently in their homes (listed in Appendix B). They could do this by sending the requested information to CSR either in an electronic file or on paper. Study staff then determined who was eligible to be interviewed.

Residents were declared ineligible if any of the following criteria were met (numbers refer to items in the MDS 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).

Once eligible persons were identified, the lists were randomized. A target number of interviews was set for each home: 25 in smaller homes (less than 115 beds), 40 in larger homes (115 or more beds). The protocol was to attempt to interview the first 25 (or 40) on the randomized list, going deeper in the list when residents were found who could not or were not willing to be interviewed, until the target number was completed.

Interviewing at the first nursing home began on June 22, 2005. The last home was completed on August 9, 2005. The data collection was done by a staff of 11 professional interviewers who work with CSR. Of those, 7 were experienced interviewers, most with many years of experience in personal interviewing, while 4 were newly hired and trained for this project. All new interviewers at CSR receive three days of

in-person training in standardized, nondirective interviewing procedures. In addition, the 11 interviewers who worked on this project then participated in a one-day training session specifically about the procedures for this study. (Training materials are in a separate document.) All the interviewers were female. They ranged in age from 21 to 75. Only one interviewer was younger than 40.

A team of interviewers was sent to each nursing home, consisting of a site coordinator and 3 to 6 interviewers. Before arriving at each nursing home, a site information form was sent to the interviewers (Appendix C). At each home, the team met with a nursing home contact person and was given a tour of the facility. Each nursing home also provided the team with a room that would serve as their home base and where the site coordinator would be stationed. Starting at about 9AM, the interviewers attempted to interview the designated eligible residents following the protocol outlined above. (See survey instrument in Appendix D1.) They worked all day, until about dinner time. Depending on how many interviews remained to be completed at the end of the day, some subset of the team returned to complete the interviewing on the next day. All interviewing was to be completed in no more than 2 days.

All interviews were done orally. Each question was read to the respondent who then could answer either verbally or by pointing to their response on a show card. Interviewers had a show card for each response task in the instrument. Show cards provided the respondents a way of seeing each response choice as well as giving them a way to answer without letting others around hear. At the end of each interview, the interviewer gave the respondent a preprinted thank you note with her name on it. (Example in Appendix E.) This note was customized for each of the 3 states in which we interviewed. The note served several functions. It served as a reminder of the interview for the respondent and as a notice to family members about what had occurred. It provided legitimacy by listing CSR's contact information, and it also listed the Medicare Customer Service phone number and the phone number for the State Medicaid office and the State Long Term Care Ombudsman's office.

At each nursing home, interviewers received a list of residents to try to interview from the CSR site coordinator. They attempted to conduct the interviews in the order they received them. When they encountered people with whom they could not complete an interview, they returned to the supervisor to get additional sample names. If they encountered someone who was asleep or occupied in some way that precluded an interview, they were instructed to return at a later time to attempt to get an interview. If a respondent was reluctant to be interviewed when first contacted, a different interviewer recontacted the person before classifying the case as a refusal.

An important part of the survey protocol was how interviewers were to decide who was cognitively 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. These were in the form of the 0 to 10 ratings that were used in the survey. Before they started the actual survey, respondents were asked the three vignette questions. In the developmental work, the ability to give meaningful answers to the vignettes appeared to be a good predictor of the ability to answer the survey questions. The Short Blessed (a test of cognitive ability) was administered at the end of the survey. The results from both of these cognitive measures were evaluated to see how well they reliably identify whether or not a person can be a respondent. 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. However, since we decided that the vignettes were not to be used as a screener, if a respondent could not answer 2 vignettes, the interviewer then asked the first non-vignette question (rating food). If the respondent could answer that, the interview continued.

After interviews were completed, they were returned to CSR, where trained data entry staff entered the answers into a computer file. All data entry was independently verified.

After the data collection was over, supervisors debriefed interviewers who worked on the project about their experiences with the procedures and their thoughts about individual survey questions. The CSR field supervisor also arranged to conduct a telephone interview with the administrator of each participating nursing home regarding the survey procedures and demands on the nursing home, to provide a basis for considering how best to design the survey protocols in the future.

PRETEST

On May 26, 2005, prior to conducting the field test, we pretested the protocol and survey. The study team consisted of 4 people - CSR's field coordinator, 2 study staff (the project manager and the research assistant), and a senior interviewer. All of the team members had taken part in the cognitive testing of this instrument over the past several years and thus were very familiar with it.

1. The Pretest Protocol

In order to test the protocol, we used one of the staff members as the "site coordinator" - who would manage the sample - and the other three would interview residents. Starting with an eligible sample of 34 residents (50.7% of the sample we started with), we were able to interview 15 people in 5 hours. We also found 3 additional cases that were ineligible (one was no longer at the nursing home, one did not speak English, and the other had a legal guardian). Six respondents were unable to complete the interview - 4 because they could not answer 3 questions in a row and 2 because of hearing difficulties. Since this was a one day protocol with no attempts at conversion, several cases were finalized with what in other homes would be considered non-final results. These included 4 (first) refusals, 3 who were sleeping, 1 who was ill, and 1 who was unresponsive. Overall, we had an eligibility rate of 46.3%. Of those eligible, we interviewed 48.4%.

2. What we learned about the protocol

- We were confident that the 2-day protocol would work. At the end of the single day of interviewing, all of the cases had been assigned and attempted at least once. If we had returned for a second day, there would have only been 9 residents to interview (those that were first refusals, sleeping, ill, or unresponsive on the first day).

- Both the morning and the afternoon were productive times to interview.

- Using a single person as "site coordinator" to manage and control the sample was workable and very efficient. Since we had a target number of interviews we wanted at each home, the tight control of sample would be important to keeping the team on track and to not releasing more sample than needed to reach the targets.

- Not wanting to overwhelm the interviewers, but wanting to provide enough sample cases that they would work efficiently, we determined that an interviewer should start with 5-6 coversheets in the morning.

- The 3 experienced interviewers had several suggestions about how to approach residents and enlist cooperation. These were written up and given to interviewers during the training.

- Before the pretest, we had agreed to the "3 Question Rule" - that when a respondent could not answer 3 questions in a row, we would stop the interview. The pretest gave us a chance to actually observe how implementing this rule would go. Because of what we learned, we decided to modify it and extend the three questions to include not being able to answer 2 vignettes and the first interview question or any subsequent 3 questions in a row. We also learned that we would need to create an "unresponsive" code - to be used when the respondent was conscious but totally disoriented or unresponsive to the interviewer.

- Finding private locations to do interviews was a challenge. And often, respondents did not want to move or go someplace else to do the interview. We realized that some of the interviews would probably have to be done in the presence of others.

- The use of the show cards was very helpful for the respondent and the interviewer. The respondent was able to both see and hear the responses, and point to an answer if desired. For the interviewer, it was a tool to help keep the respondent focused and, if needed, could increase confidentiality of answers when the interview was not done in a private location.

- We tested several different amplifiers and headphones during the pretest. In order to make this interview as accessible to as many residents as we could, we bought 2 different amplifiers and 4 different headphones to see which worked best. Although we found no difference in the amplifiers (stereo or non-stereo) we decided that the traditional headphones (that go over the head) were better for this population than the newer and smaller earbuds or over-the-ear models. Even with the amplifiers, we still had residents who were unable to participate. One had a cochlear implant that was not working (and without which the amplifier was useless) and another had a hearing aid that did not work well in conjunction with the amplifier.

3. The Pretest Instrument

The pretest version of the interview is in Appendix F. Based on our pretest experience, we modified the instrument and informed consent before the final field test. What we learned about the instrument from the pretest is described in the following section.

4. What we learned about the instrument

- The informed consent and instruction page was very long and much too complex. We edited this in order to delete unnecessary words and to use simpler words to better explain what the interview process would be like.

- The vignettes were too long and contained extraneous information. Respondents were very confused about what their task was in this section. By simplifying the instructions and the actual vignettes, we hoped this section would go smoother. We concluded that we should tell the respondents explicitly that these were made-up people and that the situations were not real and not in their nursing home.

- We continued to find problems with the question that asked about rating care received by doctors. Respondents did not know whether to include doctors only seen in the nursing home or to include those from outside. After the pretest, the question was edited to include the phrase "either here or outside the nursing home."

- In the series of questions about hearing aids, an additional question was added after the pretest to keep the series parallel to the other help questions (eye and dental care): "Since you've been in this nursing home, have you had any help from the staff in keeping your hearing aid in good working condition? "

- We found some problems with the rating of how well the nursing home met the respondent's religious needs. Despite the screener ("Is religion a part of your life?") several respondents wanted to say that the rating did not make sense for them (basically, that this had nothing to do with the nursing home). We decided not to change the question, and found that this rating question continued to be problematic in the actual field test.

- One respondent said she could not move at all, and thus felt the question about how her room was set up was inapplicable to her. In response, after the pretest, an interviewer check was created for interviewers to mark if respondents said they unable to answer because they were "unable to move without help."

- The Short Blessed, which was administered at the end of the interview, was not well received by respondents or interviewers. Interviewers felt that it was a bad way to end an interview (ending with a "test" that respondents often know they are failing). However, no one felt it should be done before the interview either. Their preference would have been not to ask it at all. This series was kept in the field test, since we knew we were going to want to compare the results of the Short Blessed to the CPS and the vignette score, as well as to the answers to the survey questions.

C. FIELD TEST RESULTS

This section describes the protocol and results from the field test of the 12 nursing homes. It does not include findings about the pretest home.

1. Sampling

Nursing homes were given the option of providing an electronic file that contained the 19 items that were needed for sampling or sending the same information on paper (either sending a printout or handwriting the information on forms we sent them). We also offered to send someone to the nursing home to help with the task of getting the needed information. Of the 12 homes in the sample, only 2 sent an electronic file with the items as requested. Two homes filled out the information on the CSR forms and 5 homes sent us computer printouts. The printouts were often confusing to read (cases spanning more than one line or page or a printout that was sorted by variable rather than by resident). Several homes sent printouts of each resident's entire MDS record (almost an inch thick for residents who had been there for a long time). Two homes arranged for CSR project staff to visit and collect the data. There was also one home that originally sent paper printouts that did not contain all of the needed information. It was decided that to facilitate the process, a CSR project staff person would go to the site to help complete the information needed.

We learned that all the homes had the information needed in electronic form. However, the majority lacked either the data processing expertise, or the staff time, to produce selected data from their files.

Of the 10 nursing homes that sent us information, none sent totally complete and correct information on their first try. CSR staff had to have discussions with all of the homes to correct problems or get additional information that was not originally provided. Most were able to get us the information (in some format) after additional contacts. Overall, for 4 of the 12 sites, CSR felt the sample collection process was very difficult and time-consuming.

CONCLUSION: We think 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 select the sample. Most nursing home staff do not have the expertise and the time necessary to assemble the needed data and execute a sampling plan. However, as is discussed below, working with the homes and carrying out the sampling requires substantial project staff time

2. Eligibility

Table 2 breaks down the findings on eligibility based on the study criteria.

The top of Table 2 presents those who were declared ineligible based on analysis of MDS records. The largest single group was those who had been rated "severely impaired in cognitive skills for daily decision making" on their last MDS assessment (n=238). The second largest group was the 170 who had a legal guardian or other legal oversight. In order to interview such people, it would have been necessary to get written approval from the guardian or overseer. This was not feasible given the time frame of the pilot, but would be possible in a more relaxed time frame. However, it is important to note that 53 of the 170 would have been ineligible for other reasons (mostly cognitive impairment).

There were also 97 people who had a discharge planned in the next 90 days and 39 people that did not meet the criteria for length of stay; they had not been in the home for at least 30 days.

In addition, to those who were deemed ineligible before the start of interviewing, interviewers found some people to be ineligible who were not identified in the analysis of the record data: 20 were deceased, 31 were discharged, 25 did not speak English, and 7 were ineligible because we found out they had a legal guardian or other legal oversight.

Table 2: In-Person Resident Survey: Sample Eligibility

Initial Listing of Residents			**1347**
Determined Ineligible Based on MDS Records[1]:			477
Has Legal Guardian/Oversight[2]		170	
Guardian	94		
Oversight	78		
Comatose		4	
Severely Impaired in Cognitive Skills for Daily Decision Making		238	
Discharge Planned		97	
Not in nursing home >30 days		39	
Found Ineligible During Data Collection:			83
Has Legal Guardian/Oversight[2]		7	
Deceased		20	
Discharged/No Longer at NH		31	
Non-English Speaking		25	
TOTAL KNOWN INELIGIBLE			**560**
Never Contacted During Data Collection:	169		
Additional Estimated Ineligible[3]:		20	
TOTAL ESTIMATED INELIGIBLE			580
TOTAL ESTIMATED ELIGIBLE:			**767**
PERCENT OF THOSE INITIALLY LISTED WHO WERE ELIGIBLE:			**56.9%**

[1] Residents could have multiple reasons for ineligibility.
[2] Those with legal guardians or other legal oversight potentially could be interviewed. The time constraints of the field test meant we did not have time to contact guardians.
[3] 11.8% of those we talked to were found to be ineligible. Using this same rate, we expect that about 20 of the 169 people we didn't talk to would also be ineligible.

A conscious decision was made not to use records for determining the ability of respondents to answer questions in English. Hence, finding some non-English speakers also was built into the design of the test. Obviously, if the full survey is implemented, the survey instrument can be translated into other languages. Of the 25 non-English speakers, most of them (n=14), spoke Spanish. Others spoke Greek, Italian, Polish, Portuguese, and various Asian languages. Finding 51 people who were deceased or discharged occurred because there was necessarily some lag between when the record data were provided for eligibility and when the data collection began. The mean number of days between CSR receiving the sample and the date of interviewing is 19.8 days. The range was from 4 days before data collection to 45 days. This was dependent on a number of things, such as the availability of nursing home staff to get us the information and their preference for which days they wanted us to interview. As discussed in more depth in the Time & Effort section, there needs to be some amount of time between getting the sample and interviewing in order to deal with missing data and variables, sampling, and creating coversheets and other materials needed for interviewing.

Finally, the table indicates that about 20 additional residents would likely have been found to be ineligible if we had contacted the 169 residents who were randomly selected not be contacted (because interviewers had adequate sample to meet their target numbers without them).

CONCLUSION: Of the total of 1347 residents in the homes' data files, 57% were eligible for the long-term resident survey. If we include those who had guardians or overseers 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. The extent to which this is a representative sample of nursing homes, in respect to eligibility rates, is unknown.

3. Data collection results

Table 3 summarizes the results of the data collection efforts. Of the 870 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. So, there were 618 residents whom interviewers attempted to interview who were part of the study population. Of those, interviews were completed with 424, which is 69% of the eligible sample that interviewers attempted to interview.

Table 3: In-Person Resident Survey: Results of Data Collection

Believed Eligible based on Records:		**870**
Found to be Ineligible:	83	
Estimated Ineligible:	20	
TOTAL ESTIMATED ELIGIBLE:		767
Never assigned for Data Collection: (Estimated Eligible)	149	
ELIGIBLE & ASSIGNED FOR DATA COLLECTION		**618**
Attempted but Not Interviewed	194	
INTERVIEWED		**424**
PERCENT OF ELIGIBLE SAMPLE INTERVIEWED		**68.6%**

Table 4 breaks down the reasons why the 194 nonrespondents were not interviewed. This table provides a lot of information about collecting data from this population.

The single biggest reason that eligible residents were not interviewed is that they were unable to answer 3 questions in a row (39% of nonrespondents, 12% of the total sample). The second largest group, 7% of the total sample, could not be roused to answer any questions at all. Thus, close to 20% of the total eligible sample, and 60% of the nonrespondents, were not able to do an interview.

Unwillingness to be interviewed was not a major source of nonresponse. Only 6% of the total sample explicitly refused to be interviewed. In addition, an additional 2% said they could not hear well enough to do an interview and 3% said they felt too ill. Interviewers felt that these three categories contain some very similar people. Some of those who said they were too ill probably did not want to be interviewed; some of those who refused did not feel up to it. These groups combined are about 10% of the total sample. (Note: Although interviewers did have amplifiers to help those with hearing loss, very few respondents wanted to use them.)

Table 4: In-Person Resident Survey: Non-Interview Breakdown

	N	% of Non-Responders (n=194)	% of Eligible Sample (n=618)
Unable to answer 3 Questions in a Row	73	38.6	11.8
Unresponsive[1]	43	22.2	7.0
Refused	39	20.1	6.3
Ill	19	9.8	3.1
Hearing Problems	15	7.7	2.4
Other	5	2.8	0.8
TOTAL	**194**	**100%**	**31.4%**

[1] Respondents were considered "Unresponsive" if they were conscious but unresponsive to the interviewer.

One of the concerns at the start of the project was that a two-day protocol might miss people who were busy. That proved not to be the case. There were only 3 people who might have been able to be interviewed but were not because a time to do the interview could not be found. (One of these residents was away visiting family, and the others were continually sleeping when the interviewer attempted to visit.)

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 an interview. For refusals, a different interviewer made the second interview attempt. In all, 8 interviews were done with residents who were initially considered "unresponsive" and 14 were done with residents who initially refused.

A total of 324 of the 424 interviews were done on the first contact with the respondent; another 80 were done on the second contact. Thus, 404 out of 424 interviews, or 95% of the interviews were done on the first or second contact with the resident.

Table 5 shows some information about the screening and interview process at the nursing home level.

Table 5: In-Person Resident Survey: Selected Results by Nursing Home

	Reported # Long Term Beds	Actual Sample Received	Ineligible (from Records)	Assumed Eligible (from records)	Found Ineligible (at NH)	Percent Known Eligible	Not Assigned	Eligible & Assigned	Refusals	Could Not Answer 3 Questions	Interviews (% of eligibles & assigned)
Nursing Home 1 (Small for Profit)	88	93	24	69	2	72.0%	26	41	7	5	25 (61.0%)
Nursing Home 2 (Small for Profit)	59	57	3	54	1	93.0%	8	45	1	11	28 (62.2%)
Nursing Home 3 (Hospital-Based)	100	105	26	79	14	61.9%	0	65	4	5	43 (66.2%)
Nursing Home 4 (Large for Profit)	101	93	26	67	12	59.1%	0	55	2	9	34 (61.8%)
Nursing Home 5 (Large Non-Profit)	200	207	81	126	6	58.0%	57	63	2	3	49 (77.8%)
Nursing Home 6 (Small Non-Profit)	63	97	50	47	12	36.1%	2	33	1	0	26 (78.8%)
Nursing Home 7 (Large for Profit)	136	146	66	80	2	53.4%	21	57	5	3	45 (78.9%)
Nursing Home 8 (Large for Profit)	92	109	41	68	8	55.0%	0	60	2	12	34 (56.7%)
Nursing Home 9 (Small for Profit)	75	92	49	43	1	45.7%	0	42	3	5	33 (78.6%)
Nursing Home 10 (Large for Profit)	120	152	39	113	6	70.4%	44	63	4	10	39 (61.9%)
Nursing Home 11 (Small Non-Profit)	59	71	30	41	3	53.5%	6	32	2	4	26 (81.3%)
Nursing Home 12 (Large for Profit)	124	125	42	83	16	53.6%	5	62	5	6	42 (67.7%)
TOTAL	1217	1347	477	870	83	58.4%	169	618	39	73	424 (68.6%)

CONCLUSIONS:

The most common reason for nonresponse is that eligible respondents were cognitively unable to answer survey questions. That accounted for about 60% of the nonresponse.

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.

Recontacting residents who cannot or will not be interviewed at initial contact is productive. However, 95% of those interviewed were interviewed on the first or second contact with an interviewer. Having a protocol that allows for more than two contacts is probably not necessary. 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.

4. Screening for ability to respond

One goal of the field test was to evaluate several systematic approaches to assessing the ability of residents to be interviewed. Those methods included short series of vignettes in which respondents were asked to rate hypothetical patient experiences, the Short Blessed, and the use of CPS scores. Table 6 shows the outcome of cases based on CPS score.

Table 6: In-Person Resident Survey: Sample Outcome by CPS score

CPS	Actual Sample Received	Ineligible (from Records)	Found Ineligible (at NH)	Percent Known Eligible	Not Assigned	Eligible & Assigned	Refused	Un-responsive	Could Not Answer 3 Questions	Interviews (% of eligible & assigned)
0	167	36	15	69.5%	17	99	4	0	7	82 (83.8%)
1	173	40	12	69.9%	23	98	4	1	7	80 (81.6%)
2	197	26	12	80.7%	31	128	10	2	10	99 (77.3%)
3	448	94	28	72.7%	72	254	16	22	45	153 (60.2%)
4	84	18	16	59.5%	12	38	5	18	4	9 (23.7%)
5	136	136	0	--	0	0	0	0	0	0
6	77	77	0	--	0	0	0	0	0	0
missing*	65	50	0	23.1%	14	1	0	0	0	1 (100%)
TOTAL	1347	477	83	58.4%	169	618	39	43	73	424 (68.6%)

* Respondents who have "missing" CPS scores are those whose records did not have all data needed for CSR to compute their scores.

NOTE: The eligibility criterion of not being "severely impaired in cognitive skills for daily decision making" eliminated those with a CPS score of 5 or 6 from the sample.

Interviewers reported that there were some respondents (at a rate hard to estimate) who could not answer the vignette questions but were able to rate their experiences. Of the 424 interviews, 26 were completed with respondents who could not answer 2 or more of the vignettes at all. Less than 40% of the respondents correctly scaled all 3 of the vignettes. Further analysis can be done to more carefully compare the vignette responses to the survey results and the other cognitive measures collected. Interviewers also reported that the administration of the Short Blessed was very stressful to some respondents (again, rate hard to estimate) because they knew they were being tested and asked reasonable questions, to which they did not know the answers. All of the interviewers said they would prefer not to have either of these measures if they are not necessary. The interviewers said that the strategy used in this field test —to stop the interview whenever a respondent was unable to answer any three questions in a row—was the least intrusive and stressful way to identify respondents who could not be interviewed.

CONCLUSION:

While eliminating only the most severely cognitively impaired (CPS scores of 5 or 6), we were able to interview some respondents with moderately high levels of cognitive impairment who might be eliminated in other protocols. Interviewers would prefer not to use a screener for cognitive ability unless it is highly predictive. Since the vignettes are not going to be used to identify respondents who can complete the interview, we recommend deleting the vignettes from the interview protocol. We recommend that everyone who meets eligibility criteria be interviewed. If a respondent is unable to answer 3 questions in a row, the interview should be stopped.

5. Data Collection Process

a. Location of interviews:

About 79% of the interviews were conducted in the respondent's room. The rest of the interviews were done in more public areas, such as the dining room, other common room, outside, or even in the hallway. The protocol requires that, if at all possible, the interview should be done in a private place. Unfortunately, this was not always the case. Only 51.2% of the interviews were done with no one else around; 42.7% of the interviews were conducted with a roommate or other resident present. For only 5 of the 207 interviews that had someone else present did the interviewer feel that the presence of someone else influenced the respondent's answers (twice interviewers felt that having another person present influenced the answers "some" and in 3 cases, they felt the respondents were influenced "a little"). Staff were not present for the entirety of any interview, though in 9 cases, they were present at some point. (In these instances, the interview was stopped until the staff member left the room.) Roommates were often asleep or engaged in other activities (such as watching TV) while interviews were going on. In common areas, interviewers were able to conduct interviews away from the crowd in as private a location as possible. Often respondents did not want to be moved from where they were to complete an interview. Interviewers made their best attempts to try to bring them to a private area before agreeing to interview in a public area or where others were present. As mentioned earlier, the use of show cards eased some privacy concerns, since respondents could point to their answer on the show card rather than saying it out loud.

b. Respondent characteristics:

Interviewers were asked to describe both the physical and emotional characteristics of the respondents. The most frequent physical characteristics are being in a wheelchair (40.1%), having a hearing impairment (9.2%), and having a visual impairment or blindness (7.8%). Interviewers reported that 6.1% of the respondents appeared to be confined to bed. Over 75% of the respondents were dressed in street clothes (rather than pajamas) when they completed the interview. When asked to assess whether the resident appeared to understand the interview questions, interviewers felt that about 66% "always" understood the questions and that another 30% understood the questions "sometimes."

CONCLUSION:

It was not easy to find a private place to administer the interview. Even for those interviews that were done with other people around, interviewers felt that it rarely interfered with the survey process. Part of this could be because of the use of show cards for the response options. As expected, many respondents who were interviewed had physical and intellectual impairments. Only about 66% of those who completed interviews were assessed by interviewers as always understanding the questions.

D. TIME AND EFFORT REQUIRED FOR CURRENT RESIDENT PROTOCOL

The estimates below do not include any travel time or costs, since those will be dependent on the location of the project staff, interviewers, and sampled nursing homes. They also do not include any information about the time required to recruit the nursing homes.

1. Sampling

We attempted to collect the sample for both the in-person (resident) and mail (discharged resident) surveys at the same time. Thus, the information provided below is for the entire sample collection process. We collected resident lists from 12 nursing homes and discharged resident data from 11 of those.

Sample collection was a very time consuming process. It involved many phone calls, emails, and in several cases actually going to the nursing home to extract data from files. Three project staff, 4 clerical staff (2 of whom were interviewers on the project), and 5 coding staff members were involved.

After the nursing home agreed to participate, the first step was for CSR project staff to talk to the designated sample person and figure out a process of getting the necessary information to us. It took about **20 hours** of effort of the process involved in making follow up calls, emails, and faxing of forms.

Two nursing homes requested that CSR help with sample collection and one home, who originally sent incomplete paper forms, needed someone to go down to help. This took about **35 hours** total for the three homes (totaling about 400 beds).

Only 2 of the nursing homes sent information in electronic format. So, for the other 10 homes, the information had to be entered into computer databases for sampling and controls purposes. For some, this was a 2-step process: first we needed to extract the data from the records and put it on a standardized form, and then enter the information from the form into the computer. This took about **37 hours**

Additional project staff time was needed to coordinate the sample entry, check the information as it arrived from the nursing homes, and make corrections and requesting additional data when problems arose. This was about **40 hours** of work.

CONCLUSION:

All together, sample collection took about *137 hours of effort*. Most of this - 72 hours - was done by interviewers and clerical level staff. The rest of the time was spent by the project manager and research assistant checking and organizing the sample data (this includes checking data, arranging for corrections, and editing data).

The cost for a larger data collection effort would depend on what percentage of homes could deliver electronic data files. If our experience is representative, we estimate that identifying a sample would average 6 hours per home clerical staff time and about half a day per home of professional staff time.

2. Data Collection

A lot of work was necessary before the interviewers arrived at the nursing home in order to ensure smooth data collection. For this study, about **100 hours** were spent on pre-study work – this includes creating the forms that the interviewers used, figuring out strategy and plans for the process, recruiting and training the 11 interviewers. This did not include any time spent on writing, testing, or formatting the instruments themselves.

Another **35 hours** were spent on scheduling the nursing homes and the interviewers required to staff the 12 nursing homes.

CSR spent about 8 hours per interviewer for 9 interviewers and 2 supervisors on the initial study training and briefing. Also, approximately 35 hours were spent by the trainers to prepare and lead the training. This is a total of 123 **hours**.

Once the interviewers were in the nursing home, in order to work through the sample and collect the 424 interviews, it took approximately **560 interviewer hours** and **155 supervisory hours**. This time includes the actual interviewing, as well as time spent looking for the respondents, convincing them to participate, and editing completed interviews. CSR's field coordinator, in addition to providing weekly supervision of all interviewers, also provided on-site supervision, observing each interviewer in the field at least twice. This added an additional **71 hours** of supervisory time.

We had originally planned that all sample would have been collected by the briefing date, since this was not the case, additional materials had to be mailed to each site supervisor prior to visiting the nursing home. This added an additional **12 hours** of clerical time. All site materials, including interviews, notes, coversheets, and other information, were sent by the site supervisor to CSR via overnight mail. Although this did not result in additional time spent on the study, this is an additional cost.

Once the interviews and coversheets were returned to CSR, it took about **32 hours** to control all of the coversheets and **35 hours** to enter and 100% verify each of the 424 interviews. An additional **55 hours** of time was necessary to write the data entry programs, extract and clean the data, and enter all of the interviewer observations from each coversheet and timesheets.

CONCLUSION: It took approximately 1178 hours of staff time to conduct this study. This number does not include the costs and time associated with travel, nor does it include time spent on basic (non-study specific) interviewer training. It also does not include the time and effort required to recruit the nursing homes. This averages to about 3 hours of staff time per completed interview. For this field test we used professional interviewers. This allowed us to be very efficient in the use of interviewing time and gave us confidence that the interview was being administered in a consistent and standardized way. Having interviewers who understood the interview process also helped make sample decisions, such as when a respondent was not able to continue the interview, a less arbitrary event. Using non-professional interviewers would change not only the cost implications, but most likely the consistency of interviewer decisions and the quality of the data they collected. It might also require additional supervisor time for non-professional interviewers.

3. Summary - Cost Estimates

Table 7 organizes the above information and puts some dollar figures on the estimates. The total direct costs for the in-person survey task that yielded 424 interviews from 12 nursing homes were just over $60,000. Of those costs, $30,000 were for the time and expenses related to travel to and from the nursing homes. This is the price that was paid for having homes an average of 2 hours from where interviewers lived and having a protocol that involved frequent overnight stays. $5850 was paid to the individual nursing homes for their help and participation. Each home was given a donation of $250 to the Resident Fund. Depending on the size of the home, an additional $150 (for small/medium homes) or $300 (for large homes) was given to help offset costs of staff time needed to prepare the sample files. Of course, this payment may or may not be part of future studies.

If one excludes travel costs, thereby including only the time when staff members were in the homes, the **direct costs for data collection were about $24,000 or about $57 per interview**. Obviously, any future survey would entail some travel, but most likely considerable less per interview than this field test design.

In addition to travel, other issues to consider when extrapolating from this experience to possible future surveys:

1. Pay rates could differ.

2. On a per interview basis, the effect of some of the starting costs might be lower for an ongoing data collection.

3. Because of the level of control we wanted to exercise, we probably had more supervision than would be necessary in an ongoing survey.

4. If the interview is shortened by dropping items or one or both of the cognitive evaluation series, costs per interview might drop a little.

5. These figures do not include indirect costs, which will be specific to the data collection organization.

Table 7: Summary of Effort and Cost by Activity

	Hours	Average Pay Rate*	Total
Sample Development (Total Cost: $2455)			
Professional	55	$25	$1375
Clerical	72	$15	$1080
Training & Briefing (Total Cost: $2720)			
Supervisory	35	$25	$875
Interviewers	123	$15	$1845
Project Organization & Management (Total Cost: $4275)			
Professional	171	$25	$4275
Data Collection at Nursing Homes (Total Cost: $12,275)			
Supervisory	155	$25	$3875
Interviewers	560	$15	$8400
Sample Control & Data Entry (Total Cost: $2380)			
Professional	55	$25	$1375
Clerical	67	$15	$1005
TOTAL WITHOUT TRAVEL			**$24,105**
Travel Expenses (Total Cost: $30,250)			
Interviewer Time	550	$15	$ 8,250
Travel Costs			$22,000
TOTAL WITH TRAVEL			**$54,355**
Payment to Nursing Homes (Total Cost: $5850)			
Donation to Resident Fund		$250/each	$3000
Payment for help with Sampling (5 small homes @ $150/ 7 large homes @ $300)			$2850
TOTAL WITH NURSING HOME PAYMENTS			**$60, 205**

* Approximate average across numerous individuals with varying rates.

We also calculated costs per home and per interview for large and small homes. There were costs, such as interviewer training, sampling and setting up the home, that were unaffected by home size. These fixed costs came out to about $790 per home for 12 homes. The data collection costs per home were basically proportionate to the number of interviews: At **about $30 per interview for interviewing and supervision**, small homes cost about $846 for an average of 28 interviews and large homes cost about $1150 per home for an average of 41 interviews. Sample control and data entry were also proportionate to the number of interviews. The costs were about $5.60 per interview. This came out to about $155 for each small home and $230 for each large one. All of these estimates exclude travel and overhead.

E. FEEDBACK ABOUT CURRENT RESIDENT PROTOCOL

1. Feedback from Nursing Home Administrators on Protocol

After the field period, the CSR field supervisor called each nursing home administrator/contact person to discuss how the process went for them (see Appendix G for questions discussed). She was able to talk to 10 of the 12 homes (though at one home, she was not able to complete the interview). For one nursing home, we have information from 2 different people - the MDS coordinator and a MDS nurse, whose answers vary somewhat.

a. The Sample

How long did it take to collect the sample? It took an average of 8 hours for each nursing home 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. One nursing home said it took 1-2 hours to collect the information, and 3 nursing homes said it took 3 or more days of someone's time. 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. These numbers include the time it took to check information, reply to calls from CSR about problems, and, sometimes, resend the data.

Everyone seemed a bit surprised about the amount of work it took to get the sample together. Although they said the CMS and Harvard introductory letters were fine, they did not seem to understand the effort it would take to get the data we needed. The time commitment on an already overly committed staff was unanticipated.

Were all necessary data in one place? Although everyone had some files on computer, 6 sites had to access paper records for at least some information. (Two of these 6 sites had to access paper files only for information about guardians and conservators.) Several people told us that they had to access several computer systems (including business office, MDS, and quality assurance files) to get us all of the variables we requested. Even those who had everything in one database still had to access both the most recent MDS assessment, as well as the admission assessment (for items including type of payment, age, education, and guardian status). This definitely increased the burden on the nursing home. The hardest piece of information for people to give us was whether the respondent had a guardian or conservator.

Was there a difference getting current resident vs. discharged resident data? It was harder for nursing homes to give us the data for the discharged sample. Five of the 10 nursing homes said that discharged files were not in the same location and required extra work to process - often requiring having to access paper files. Information about where the resident was discharged to and to whom did not seem to be in the same files as the other information and often came to us in different files and on paper. Since all nursing homes do not store data in the same format, figuring out how to gather what we needed was complex. For at least one home, where the records are stored is dependent on the discharge date (how long ago) and where the resident was discharged to (information about deceased residents not always filed with other types of discharges). In another home, how one accessed files was dependent on the length of time in the nursing home before discharge - sometimes the full MDS assessment was not in the computer files yet and had to be retrieved from the paper copies.

Overall feeling of how sampling went: At 9 of the 10 nursing homes, the administrators felt the sampling process went well. The only comments were about added time it took from their already busy job to get us the information. It did not go well for one site. At that nursing home, the MDS coordinator was in the process of leaving and was already very busy transferring her duties to someone else. She was not happy the administrator asked her to do this and had trouble sending us the information we requested. After sending files with some of the data (and sample) missing, and after numerous phone contacts, it was

finally agreed that CSR would send staff to collect the data.

Several people asked about why certain pieces of information were requested. This is a legitimate question, since in several situations; it caused an increase time burden on the nursing home.
Of the 19 items requested:

- 8 were vital for creating the sample: name; room number; AB1 (date of admission), B1 (comatose), B4 (cognitive skills), Q1c (whether discharge was planned), and A9a & A9b (have legal guardian or legal oversight);

- 3 items were helpful in the interviewing and sampling process but not necessary to have: Patient NH identifier, AA2 (gender), and AA3 (birth date)

- In addition to the items needed for sampling, an additional 3 were needed to create the CPS score: B2a (short-term memory problems), C4 (making self understood), and G1hA (eating self-performance)

- 5 items were of interest for analysis during this field test (and most likely would be in any future testing of this protocol: AA4 (race/ethnicity), AB7 (education), A3 (date of most recent MDS assessment), and A7a & b (Current Payment Source: Medicaid or Medicare)

Of these, we feel that only the payment source items (A7a and A7b) definitely could be eliminated from requests during further testing.

Two sites asked about why we did not access data from the state database or CMS records. They said that when state comes in they know who they are looking for and that CMS has all the quality and MDS data. One of these sites also wondered why we didn't use the method state surveyors do - using staff judgment about who is "interviewable and reliable."

Could the nursing home have handled sampling? When asked, most said they could have created lists for us of residents who met certain sampling criteria. However, we find this quite unlikely. Considering the problems of getting simple census data from these sites, it is hard to imagine they could do the sampling required, even with a lot of help. Perhaps if the sampling criterion was very simple (such as everyone admitted after a certain date), it would be workable. However, considering our inclusion and exclusion rules, it would be an additional burden on the nursing home and could be done incorrectly.

b. The Days of Interviewing

All of the nursing homes thought the actual interviewing process went well. Administrators had high praise for the interviewing team, their friendliness, and their self-sufficiency. On the whole, there were no disruptions or difficulties. At most of the homes, there seems to have been some preparation before the staff arrived. One home sent out a letter to family members of current residents explaining what would be happening. Another home specifically mentioned talking to the Resident Council about us. We don't know how many had briefed their staffs, though several did mention that they did. However, a couple of administrators said they got questions from the staff about who we were and why we were there, leading us to think the information was not as well disseminated as we would have hoped.

Of the hundreds of residents spoken to, only 2 (at two different homes) complained to staff. Both administrators who told us about this felt the situations were isolated instances. In one situation, a resident complained that the interviewer was still in the room when it was lunchtime (though, the lunch trays had not been delivered to that floor yet). In the other case, a resident wanted to stop near the end of the interview and interviewer tried to get her to finish.

Administrators at three nursing homes mentioned that they would have liked to see the sample list before we started interviewing. For one, the intent was to help us find the residents. It was felt that if they knew who we were interviewing, the staff could get them ready for us. The other two administrators wanted to provide input about who we interviewed. One administrator felt that before beginning interviewing, the team should talk to Director of Nursing about the residents to see if anything that day could be affecting their cognitive abilities, whether they had any emotional issues, or if the resident was capable of doing an interview. Another wanted to see the sample list beforehand so the Social Services team would be able to explain to some selected residents who we are and what we are doing. He gave the example of a psychotic resident who would very likely think we were from the FBI if we started asking him questions.

The administrator did not want to talk to everyone sampled, just certain people he was concerned about. The concerns of this last nursing home were brought up before we interviewed. The procedure we came up with allowed him to give us a list the morning we arrived of those patients he felt we shouldn't talk to (without prior preparation) and we would not interview them. There was only one person who was removed from the sample because of this problem.

c. Other surveys

Five of 9 sites do some sort of current resident satisfaction survey. Four of them do it yearly, and one does it every 6 months. Three of them specifically stated that it was sent to the "responsible party." At one of these homes, the vendor comes into the nursing home and hands out forms for the residents to fill out themselves. Seven of 9 nursing homes send satisfaction surveys to discharged residents. Four of these homes send the surveys within 30 days of discharge.

d. Overall feelings about the survey protocol

We asked people whether they would prefer an interview process that utilized a team of interviewers over 2 days or if they would rather have a couple of interviewers come in on their own over a longer period (like a week). About half of the administrators we talked to said they preferred a team of interviewers over 2 days rather than fewer interviewers over a longer period. Everyone had the same basic reasons – there was a lack of space at the nursing home (both in terms of parking and office space to do interviews) and they would prefer to get it over with as quickly as possible. Most of the others did not have a preference for either one way or the other.

2. Feedback from Interviewers on Protocol

We asked the interviewers and the on-site coordinators how they felt about the interviewing process overall.

- Interviewers thought the process went very well and enjoyed doing the interviews. Usually, field interviewing is a solitary process. Interviewers enjoyed the chance to work in a team and liked working on the study.

- Interviewers thought that having a "team" with an on-site coordinator was the most efficient protocol. They felt without that, it would have been more of a burden on the nursing home staff.

- The training process (how to work with an elder population) and having nursing home specific information (such as parking, contact names, directions) was helpful.

- We should make sure the administrator alerts the staff to the fact interviewing would be occurring.

- Using show cards worked well.

- Having other people present during the interview did not seem to matter to residents - almost no one paid attention to them (especially to roommates).

- Some number of respondents (maybe 10%) did not believe that information would not be given to staff and had concerns about confidentiality.

- Amplifiers weren't used a lot but having them was very helpful.

- No one liked the Short Blessed - it was often awkward to administer and was hard to record answers. Some felt that it also unnecessarily increased the length of the interview. Many also commented that they felt it was a bad way to end the interaction.

3. Feedback from Interviewers on Interview Schedule

Interviews felt the length of the actual instrument was fine. This section highlights some of the questions the interviewers thought were most problematic.

- Vignettes

Interviewers who participated in the cognitive testing and the pretest felt that this version of the vignettes was much better and much easier for respondents to answer. However, they still felt many respondents were confused by the fact these are hypothetical situations. Respondents would ask who this person was and felt they could not answer because they had no personal knowledge. We should consider being more explicit that these are made-up people and are not at their own nursing home.

- Question 6 (safe and secure, C6)

Not sure if everyone understood it consistently.

- Question 10 (how quickly staff come when called, C10)
A few people first answered that they "don't call for help." Also, there was some feeling that a few respondents were concerned about the confidentiality of their responses to this question.

- Question 15 (how clearly staff explain about care, C15)
There was some concern that respondents weren't listening to the entire question and answering just about how well staff explain things (not necessarily about their care).

- Questions 17 & 18 (seen doctors, rate care from doctor, C17 & C18)
Everyone agreed that these were problematic questions. There was lots of discussion here between the interviewers and the respondents. They felt respondents were not hearing "either here or outside the nursing home." Many respondents said things like "I have my own doctor" and "I see him at his office." Also, several interviewers mentioned that some respondents were not focusing on getting "care," but whether they "saw" someone. When these respondents got to Q18, they felt they didn't get any "care" so they couldn't rate it (for example, one respondent answered "I didn't get care, someone just came in to talk to me and take my pulse.")

- Questions 19-21 (hearing aid series, I_q19 – I_q21)
Respondents sometimes answered that they have a hearing aid but don't use it, so they skip out of the series. Some respondents were not including changing batteries as "help," and some even wondered if getting help locating a lost hearing aid should count. A few respondents mentioned that they didn't feel it was the responsibility of the nursing home to do this.

- Questions 22-25 (eye and dental care, I_q22 – I_q25)
These questions seemed to work better than the hearing aid ones. However, it was not clear if respondents are seeing this as inclusive as it was intended. Some interviewers thought that some answers to the screeners were false negatives - that residents actually got help but said no. Also for the dental questions, some respondents immediately answered "I have dentures" and didn't listen to the whole question (even though dentures are mentioned in the list of things to include.)

- Questions 26 & 27 (religion part of life, rate help with religious needs, C22)
Interviewers all agreed that Q26 was extremely problematic and felt that it was probably the most

repeated question in the entire survey. Several also felt that this was an abrupt change of pace in terms of question placement. Some respondents commented they did not know how to answer because religion once was a part of their life but is not now. Overall, Q26 is not a good screener for Q27. Many respondents who felt that religion was important to them felt it had nothing to do with the nursing home and that it was a personal matter. At Q27, many respondents simply told us what kinds of services the nursing home has. They didn't connect the nursing home and what they expect from the home to their idea of having religious needs met.

- Question 31 (place to visit in private, C27)

This seemed to work well, although some felt their own room was the private place (even when they had a roommate present).

- Question 32 (how room was set up, C28)

This was another difficult question for both interviewers and respondents. Interviewers felt uncertain about the added response category for those respondents who could not move without help. Like a tailored inapplicable option in a phone survey, the respondent did not know that option existed, and thus did not know they could say that as an answer. Interviewers felt that if it was obvious that the respondent could not move on their own, asking the question seemed unnecessary. Respondents were also confused by this question. Interviewers reported false positives here - respondents said "yes" though the room was obviously not set up to meet the resident's needs. For some respondents, the first answer was often "I can't move around much" - rather than answering about the room set-up.

- Questions 37- 39 (choice, C33-35)

Some interviewers suggested that some people didn't know what to do if they didn't do it, so they suggested adding a phrase that said "if you wanted, could you choose . . ."

- Question 40 (asking about things to do, C36)

Interviewers felt that this question was answered by most people as if it was asking about the daily activities and what was on the calendar to do. This was about whether anyone came around to say "we're having balloon volleyball in 10 minutes." A few respondents clearly indicated they were talking about what happened at intake when someone asked what they liked to do. Also, some interviewers mentioned that there were some false negatives given by respondents who opted out of activities, so they said no

(because they didn't choose to do any, rather than answering about choice of activity or choice of doing them or not.) This question does not seem to measure what we intended.

- Question 47 (feeling about life now, C44)

This question received mixed reviews by interviewers. Some felt that it was an awkward question to ask and that it was a hard question for respondents to answer. Others felt that even though it was hard, most respondents were thoughtful about their answer and took some time before responding. There were cases of respondents getting emotional here - but no major problems overall.

CONCLUSION:

The real problem questions for deletion or major revision are:

Q17 - 18 (doctor care)

Q20 -21 (help with hearing aid)

Q26-27 (religious support)

Q32 (room set-up)

Q40 (asking about things to do)

Q47 (feeling about life now) is OK, but it would be good to drop it if it doesn't play an important role in analysis.

As we mentioned earlier, we also recommend deleting the vignettes.

5. Length of Interview Schedule

Table 8 presents the mean and ranges of time for the separate parts of the instrument. The questionnaire itself seems to be about the right length. About 83% of the interviews took 20 minutes or less to complete (not including the vignettes or the Short Blessed). There were only 15 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.

Table 8: Interview Length by Section

	Mean Time (minutes)	Range (minutes)
Vignettes	3.29	1 - 12
Survey	15.94	5 - 80
Short Blessed	4.41	1 - 15
Total	23.66	10 - 84

II. MAIL SURVEY OF SHORT-TERM DISCHARGED RESIDENTS

A. BACKGROUND

Short-term nursing home residents are a very different population than long-term residents. By definition, their stays are usually to deal with an acute medical need. Their contact with the nursing home is primarily a medical event, not a total living experience. Most importantly, the cognitive problems that are so prevalent among long-term residents are much less prevalent in the short-stay population.

For these reasons, and because short-stay residents of nursing homes are geographically scattered after discharge, it was decided that a mail survey was the appropriate mode to collect data from this population. For convenience and timing, the short-stay residents of the nursing homes participating in the long-term field test were used as the sample frame for this effort.

The survey instrument for long-term residents was adapted slightly (adding a question about therapy received and deleting questions about help with hearing aids, and eye and dental care) and put into a mail survey form (Appendix D2). Since the methodology for doing mail surveys with Medicare populations is well established, the main motivation for this phase of the pilot study was to learn about eligibility and collect some data from this population that could be compared with the results from the survey of long-term residents.

This report will cover the procedures and data collection results.

B. PURPOSES

1. To learn how many people would be eligible for this survey from the sampled nursing homes.
2. To learn about the quality of information available from nursing homes for sampling and contacting this population.
3. To get some experience with the willingness of this sample to respond to a mail survey.
4. To collect data to compare with the results from the long-term resident survey.

C. DESIGN & PROCEDURES

At the same time the nursing homes participating in the long-term resident survey were providing data about their long-terms residents, they were also asked to provide a list of all those who had been discharged from the nursing home within the past 2 months. As with the current residents, we asked for several pieces of information from their MDS files (Appendix H).

Residents were declared ineligible if any of the following criteria were met:

1. Those who had a guardian or other legal oversight (A9a = 1 or A9b = 1).

2. Those whose last MDS evaluation indicated they were "severely impaired in their cognitive skills for daily decision making" (B4 = 3).

3. Those who had not been in the home for at least 5 days or who had been in the home more than 90 days (AB1 < 5 days or > 90 days from discharge date).

4. Those who were deceased or discharged to another care facility (R3a not 1, 2, or 3).

The protocol was to send an initial mailing, with a cover letter, fact sheet, copy of a self-administered questionnaire, and a postage-paid return envelope (Appendices I & J). A packet with a second questionnaire was to be sent after two weeks to the non-responders (Appendix K). Then, nonresponders were to get a reminder call to make sure they had received the questionnaire, answer any questions, and urge participation.

The plan was for the initial mailing to occur in mid-June. However, because nursing homes were not recruited and sample was not delivered until later, the first mailing was postponed until June 28. A major challenge for this phase of the study resulted from the fact that 5 nursing homes could not provide their completed samples until well into July. Because of a hard deadline for data collection of mid-August, there was not time for follow-up mailings to those nursing homes. Those homes received their first (and only) mailing at the same time as the other 6 received their second mailing (July 14). One nursing home was not recruited until the end of July. Although in-person interviews were collected, this home did not participate in the mail portion of the pilot test.

Questionnaires were returned in postage-paid envelopes to the Center for Survey Research, where the data were entered and 100% verified.

Starting July 28, CSR called everyone who had not yet returned a completed questionnaire. The purpose of the call was to prompt non-responders to return the written survey. Telephone reminder calls were conducted by the professional interviewing staff at the Center for Survey Research, from a central interviewing facility on the UMass Boston campus using a computer-assisted telephone interviewing (CATI) format. Again, because of time constraints, a limited number of calls were attempted. Each case had up to 3 calls made on different days and at different times of the day to attempt to make contact with the appropriate respondent.

D. FINDINGS

1. Sampling & Eligibility

The sample for the discharged residents was requested at the same time as the current residents. The information needed for this part of the project included the date of discharge as well as the location and phone number of where the resident was discharged. For some nursing homes, this was an additional burden since the discharge information was not always in the same location as the current residents' medical records. For discharged residents, nursing homes were asked to send information about all residents who were discharged in the last 2 months. For some homes, this meant the previous 2 calendar months and for others within 60 days of the date the created the file.

Table 9 presents the findings on eligibility based on the study criteria.
The initial sample consisted of 381 patients from 11 nursing homes. Thus, the nursing homes that had over 1300 long-term residents had about a third that many short-stay patients in the 60-day reference period.

Table 9: Mail Discharged (Short Term) Resident Survey: Sample Eligibility

Initial Listing of Discharged Residents			**381**
Determined Ineligible Based on Records[1]:		133	
Has Legal Guardian/Oversight[2]	11		
Guardian	5		
Oversight	6		
Discharged to Another Care Facility	60		
Deceased	36		
In nursing home > 90 days (alive)	12		
Not in nursing home >5 days (alive)	17		
Severely Impaired in Cognitive Skills for Daily Decision Making	7		
Total Mailed			**248**
Found Ineligible During Data Collection:		10	
Deceased	10		
TOTAL INELIGIBLE		143	
TOTAL ELIGIBLE:			**238**
PERCENT OF THOSE INITIALLY LISTED WHO WERE ELIGIBLE:		**62.5%**	

[1] Residents could have multiple reasons for ineligibility.

Almost half of the 133 residents who were ineligible were so because they were discharged to another care facility and not discharged home. Since those who died in the nursing home were also listed as "discharged," another 36 cases were excluded on that criterion. In 29 cases, the resident did not meet the residency guidelines of being in the nursing home at least 5 days but not more than 90 days. There were very few cases (n=10) who were eliminated because of having a legal guardian or other legal oversight or being severely cognitively impaired (n=7). Ten other cases were found to be deceased based on mail returns and information gathered during the telephone follow-up calls.

CONCLUSION: Eleven homes provided us with information about 381 residents who had recently been discharged. In future studies, clearer guidelines should be provided about what exactly "2 months" should include. Slightly more of the discharged residents were eligible than the current residents. The major reason for ineligibility was being discharged to another care facility or being deceased. Very few in the sample were ineligible because of cognitive impairment or having a legal guardian/legal oversight. As with the long term residents, the extent to which this is a representative sample of nursing homes, in respect to sample size or eligibility rates, is unknown. In order to obtain sufficient levels of response for reporting results by individual nursing home, future research might consider different sampling options, including using a rolling sample, where residents over several months are surveyed, rather than just the 2 month window that was used for the field test.

2. Data Collection

Table 10 summarizes the returns from the mailing.

We found that the quality of the contact information was generally fairly good. About 70% of the sample sent from the nursing homes included phone numbers. One nursing home did not provide any phone numbers, which accounts for about 40% of the missing phone numbers. Telephone reminder calls were attempted for all mail nonresponders. Two weeks after the 2nd mailing, 148 cases were sent to the phone room. There were only 30 cases, or 20% of the nonrespondents, for which we were not able to make telephone contact with either the former resident or someone else in their household - and there were 11 other households where we thought we had the right phone number, though we were never able to talk with anyone. Of the 248 total mailed sample, this means we could not contact only about 17% of the households. Although losing 10 to 20% of a sample due to poor contact information is a problem, that is probably a better than average experience with this kind of list sample information. It means 80% at least will have a chance to respond.

As Table 10 shows, only 52% of the sample thought to be eligible actually returned a questionnaire. However, five of the 11 nursing homes provided sample so late that discharged residents only had one mailing and had a very truncated period in which to respond. If one calculates the response separately for homes that received one vs. two mailings, the one-mailing homes averaged 43%, while the two-mailing homes averaged about 57%. Thus, there is reason to think that if a standard mail protocol with telephone

follow-up had been used, response rates would have exceeded 60%. There were only 9 explicit refusals, either through mail or phone contact. That is less than 5% of the total sample. Given the number of respondents and family members that interviewers were able to contact by phone, if the protocol had included the option of having the interviewer conduct a phone interview, certainly the response rates would have been much higher.

The level of cooperation is particularly interesting in that almost all of the nursing homes we were working with routinely conduct their own survey of discharged patients. Some respondents no doubt received both surveys, but there was no apparent adverse effect on response.

Table 10: Mail Discharged (Short Term) Resident Survey: Results of Data Collection

Believed Eligible based on Records (Total Mailed):			**248**
NH with 1 mailing	98		
NH with 2 mailings	150		
Found to be Ineligible:		10	
NH with 1 mailing	3		
NH with 2 mailings	7		
TOTAL ELIGIBLE:			**238**
NH with 1 mailing	95		
NH with 2 mailings	143		
Non-Interviews:		**125**	
NH with 1 mailing	54		
NH with 2 mailings	61		
RETURNED SURVEYS			**123**
NH with 1 mailing	41		
NH with 2 mailings	82		
PERCENT OF ELIGIBLE SAMPLE COMPLETING SURVEY			**51.7%**
NH with 1 mailing	43.2%		
NH with 2 mailings	57.3%		

Table 11 shows some information about the survey process by nursing home and Table 12 by CPS score.

Table 11: Mail Discharged (Short Term) Resident Survey: Selected Results by Nursing Home

	Reported # Short Term Beds	Sample Received	Ineligible (from Records)	Assumed Eligible (# Mailed)	Found Ineligible (Deceased)	Total Eligible (% of Sample)	Refusals	Completed Questionnaires (% of eligibles)
RECEIVED ONE MAILING								
Nursing Home 3 (Hospital-Based)	25	13	3	10	0	10 (76.9%)	0	3 (30.0%)
Nursing Home 4 (Large for Profit)	14	42	21	21	0	21 (50%)	1	6 (28.6%)
Nursing Home 10 (Large for Profit)	60	54	22	32	0	32 (59.3%)	0	20 (62.5%)
Nursing Home 11 (Small Non-Profit)	15	29	11	18	1	17 (58.6%)	2	7 (41.2%)
Nursing Home 12 (Large for Profit)	26	41	24	17	2	15 (35.6%)	0	5 (33.3%)
RECEIVED TWO MAILINGS								
Nursing Home 1 (Small for Profit)	12	28	6	22	1	21 (75%)	1	11 (52.4%)
Nursing Home 2 (Small for Profit)	20	39	17	22	1	21 (53.8%)	0	13 (61.9%)
Nursing Home 5 (Large Non-Profit)	20	21	10	11	1	10 (47.6%)	0	4 (40.0%)
Nursing Home 6 (Small Non-Profit)	39	58	9	49	2	47 (81.0%)	4	31 (66.0%)
Nursing Home 7 (Large for Profit)	20	14	4	10	2	8 (57.1%)	0	3 (37.5%)
Nursing Home 8 (Large for Profit)	31	42	6	36	0	36 (85.7%)	1	20 (55.6%)
TOTAL		381	133	248	10	238 (62.5%)	9	123 (51.7%)

Table 12: Mail Discharged (Short Term) Resident Survey: Sample Outcome by CPS score

CPS	Actual Sample Received	Ineligible (from Records)	Assumed Eligible (# mailed)	Found Ineligible (Deceased)	Eligible (% of Sample)	Refused	Completed Questionnaires (% of eligibles)
0	158	34	124	3	121 (76.6%)	3	70 (57.9%)
1	56	16	40	1	39 (69.6)	1	20 (51.3%)
2	67	32	35	2	33 (49.3%)	1	16 (48.5%)
3	63	28	35	4	31 (49.2%)	2	11 (35.5%)
4	9	6	3	0	3 (33.3%)	2	1 (33.3%)
5	3	3	0	--	--	0	0
6	2	2	0	--	--	0	0
missing*	23	12	11	0	11 (47.8%)	0	5 (45.5%)
TOTAL	381	133	248	10	238 (62.5%)	9	123 (51.7%)

* Respondents who have "Missing" CPS scores are those who did not have all data in all of the variables needed to complete their score.

NOTE: The eligibility criterion of not being "severely impaired in cognitive skills for daily decision making" eliminated those with a CPS score of 5 or 6 from the sample.

CONCLUSION: Overall, almost 52% of the eligible sample returned a completed survey. With time to implement a good mail protocol (including an option of telephone interview), the field test experience would suggest acceptable rates of return can be achieved. In the field test, the discharged residents who received the more standard 2-mailings had a 57% response rate, while those that only received one mailing had a 43% response rate. There were very few explicit refusals to participate. The difference in results by CPS score could be influenced by the small number of cases and the shortened protocol for a third of the sample.

C. CSR TIME AND EFFORT (Discharged Resident Protocol)

1. Sampling

Since information about the sample for both the in-person (current resident) and mail (discharged resident) survey were collected at the same time, there are no discrete costs separated here.

2. Data Collection

Mailing: It took approximately **10 clerical hours** to do the mailings associated with this protocol. It also took about **4 hours of professional time** to put together the letters and labels for the survey and to spot check the mailings.

Coding: Once the questionnaires were returned to CSR, it took about **5 hours** to control the instruments and **12 hours** to enter and 100% verify each of the 123 surveys. Since the data entry programs for the current and discharged residents were extremely similar, it took much less professional time. An additional **20 hours** of time was necessary for these tasks.

Phone Reminders: It took another **20 hours** to program the reminder calls and to manage that part of the study. It took approximately **28 hours of time in the phone room** by interviewers and supervisors to complete this part of the task.

CONCLUSION:

The main purpose of this part of the pilot test was not to test the feasibility or cost of doing a mail survey, but to collect data from this population that could be compared to data collected in interviews with long-term residents. We learned, however, that getting the sample of discharged residents from the nursing homes was not as easy as getting a sample of current residents. In the study homes, there were many fewer short-term than long-term residents. The quality of the contact information provided by the nursing home was reasonable. We do not know how much of the apparent difficulty of getting sample information was the result of asking for additional information that was required for sampling and analytical purposes, beyond simply asking for the name and address of the resident that are required in order to execute the survey at all.

Most nursing homes currently administer their own post-discharge survey. However, this did not stop respondents from participating in this one. Indeed, while the time constraints and decision not do to telephone interviews of mail nonrespondents limited the response rates in this field test, given a proper dual-mode protocol, including the option of a phone interview, it is likely that we could achieve a very good response rate on this survey.

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

III. Analyses of Survey Data

In this section of the report, we describe analyses of survey data collected. Appendix L presents a crosswalk between the question numbers for comparable items in the mail survey and interview, as well as the analytic variable name for the combined data.

Prior to analyzing the data, all data were checked for ineligible codes and discrepancies resolved. Next, we conducted several types of analyses to check the data.

Analyses of completeness and integrity of data

We had data on some sociodemographic characteristics of residents from both the Nursing Home Minimum Data Set (MDS) data and the survey. Table 13 presents information about the concordance between these two sources of information.

Table 13. Comparison of demographic information from the survey and MDS data

Age/Year of Birth

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Match	430	75.97	430	76.84
Inconsistent	26	4.59	456	80.42
Only MDS	109	19.26	565	80.60
Only Survey	1	0.18	566	100.00

Race

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Match	515	90.99	515	91.01
Inconsistent	11	1.94	526	92.95
Only MDS	11	1.94	537	94.89
Only Survey	29	5.12	566	100.00

Gender

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Match	561	99.12	561	98.41
Only MDS	5	0.88	566	100.00

Education

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Match	259	45.76	259	45.50
Inconsistent	153	27.03	412	72.84
Only MDS	19	3.35	431	76.19
Only Survey	135	23.81	566	100.00

To learn about possible reasons for the observed discrepancies, we re-reviewed the original data forms for cases with gender discrepancies. Respondent names and interviewer notes were used to clarify and correct items. We determined that the MDS data were incorrect for gender for 2 cases and the remaining two discrepancies were due to ID mismatches. These errors were corrected prior to conducting other analyses.

To assess the extent to which age discrepancies may have been due to variability in cognitive functioning, we compared cognitive ability scores for those with different degrees of concordance on the age variable (Table 14). We used three measures of cognitive ability, the Cognitive Performance Scale (CPS) (0= not impaired, 6 = severely impaired) from the MDS, the Short Blessed (0= not impaired, 28=severely impaired) administered during the interview, and vignette scores which count the number of correctly ordered pairs (In some earlier analyses, a weight of 2.0 was applied to each incorrect pair. For the work reported here, we scored correct pairs so a score of 0=none correct, 1= one correct; 2= two correct; 3=perfect score.). Using the CPS score and the Short Blessed scale, we found that respondents with age discrepancies had more cognitive impairment. The relationship was less clear for vignette scores, though those with more than five years age discrepancy had lower vignette scores (Table 14).

Education discrepancies appear to result largely from differences in the response categories between the items.

Table 14. Comparison of respondents with different degrees of discrepancy between self-reported and MDS age, in terms of cognitive ability

CPS Score

Age Discrepancy	N	Mean	Standard Deviation	Minimum	Maximum
Match	430	1.54	1.19	0.00	4.00
1-4 Years	15	2.40	1.12	0.00	4.00
>=5 Years	11	2.55	0.93	0.00	3.00

Short Blessed Score

Age Discrepancy	N	Mean	Standard Deviation	Minimum	Maximum
Match	430	12.54	8.19	0.00	28.00
1-4 Years	15	16.64	8.34	0.00	26.00
>=5 Years	11	17.10	8.23	4.00	28.00

Vignette Score

Age Discrepancy	N	Variable	Mean	Standard Deviation	Minimum	Maximum
Match	430	Vignette Score	1.93	1.10	0.00	3.00
1-4 Years	15	Vignette Score	2.07	1.00	0.00	3.00
>=5 Years	11	Vignette Score	1.30	1.16	0.00	3.00

Conclusions:

1) Having multiple sources of information about sociodemographic characteristics can help identify mismatched records and MDS data errors.

2) Multiple sources of information are more useful when response categories are identical. MDS categories for some demographic characteristics should be considered for the CAHPS Nursing Home surveys so that data can be compared more easily. We do not recommend dropping questions from the survey because of the possibility that MDS data are incorrect and the possibility that MDS data will not be available in certain contexts.

Comparisons of interview respondents, interview non-respondents, mail respondents, and mail non-respondents

Table 15 presents selected characteristics of responders and non-responders in both the interview and mail samples. The potentially eligible but not interviewed residents (many were unassigned to interviewers, others failed to complete the interview) were similar to respondents with respect to gender, race, education, and payer mix (Table 15). However, those not interviewed were older (average age of 84 vs. 81years old) and more cognitively impaired. There were no significant differences between responders and non-responders in the mail sample. When we compared the interview sample and the mail responders, there were fewer African American mail respondents. The mail responders were significantly less cognitively impaired, and somewhat less likely to report feeling worried than the interview sample. There were small marginally significant differences in self-reported health status with mail respondents reporting somewhat worse health (Table 15).

Table 15. Characteristics by and Across Samples (mail and interview) of Respondents and Non-Respondents

		In-person survey			Mail survey			p value interviewed vs. mail respondents
	Overall	Interviewed	Not Interviewed	p value	Respondents	Non Respondents	p value	
Number of subjects	1063	439	379		127	118		
Age - mean (std)	81.7 (11.1)	81.4 (10.6)	84.3 (11.3)	<.001	79.5 (9.1)	77.2 (11.7)	.09	.059
Gender - Male (%)	28.1%	27.1%	25.1%	.52	31.3%	38.1%	.22	.43
Race (%)				.32			.45	.012
1.American Indian	0%							
2.Asian	0.4%	0%	0.8%		0%	0.9%		
3.African American	7.5%	8.8%	9.1%		1.6%	4.2%		
4.Hispanic	1.8%	2.2%	1.9%		0.8%	0.9%		
5.White	90.4%	89.1%	88.2%		97.6%	94.1%		
Education (%)				.14			.70	.15
1.None	1.5%	1.2%	2.9%		0%	0%		
2.8th or less	19.5%	17.5%	25.7%		12.4%	15.4%		
3.9-11 grades	12.2%	14.0%	12.7%		4.5%	11.0%		
4.High School	39.7%	39.8%	33.0%		51.7%	48.4%		
5.Tech-Trade School	7.6%	8.0%	6.5%		9.0%	7.7%		
6.Some College	9.3%	9.2%	9.8%		9.0%	8.8%		
7.Bachelors	6.9%	7.7%	5.8%		9.0%	5.5%		
8.Graduate degree	3.3%	2.7%	3.6%		4.5%	3.3%		
Payor Mix (%)				.48				
1. Both	3.8%	2.9%	4.7%					
2. Medicaid	65.3%	66.1%	64.4%					
3. Medicare	10.9%	11.7%	10.0%					
4. Neither	20.0%	19.3%	20.9%					
CPS score - mean (std)	1.87 (1.3)	1.81 (1.2)	2.42 (1.1)	<.001	0.59 (1.0)	0.78 (1.1)	.19	<.001
0	26.2%	19.6%	9.3%	<.001	71.4%	60.0%	.22	<.001
1	14.5%	18.7%	11.8%		7.6%	14.6%		
2	19.6%	24.7%	18.1%		12.6%	12.7%		
3	34.6%	34.9%	49.6%		7.6%	12.7%		
4	4.9%	2.1%	11.2%		0.8%	0%		
Rating of Health - mean (std)	2.92 (1.1)	2.87 (1.07)			3.08 (1.02)			.052
1. Excellent	9.7%	11.0%			5.0%			.15
2. Very good	24.8%	24.7%			25.6%			
3. Good	16.1%	36.9%			33.1%			
4. Fair	22.8%	21.1%			28.9%			
5. Poor	6.6%	6.3%			7.4%			
Feel worried - mean (std)	2.53 (1.1)	2.44 (1.1)			2.88 (1.0)			<.001
1.Often	18.2%	20.1%			10.8%			<.001
2. Sometimes	34.6%	38.6%			20.8%			
3. Rarely	22.8%	18.5%			38.3%			
4. Never	24.4%	22.9%			30.0%			
Feel happy - mean (std)	1.76 (0.9)	1.77 (0.9)			1.70 (0.9)			.43
1.Often	46.1%	44.7%			51.7%			.46
2.Sometimes	38.2%	39.8%			32.2%			
3. Rarely	9.2%	8.9%			10.2%			
4. Never	6.4%	6.6%			5.9%			
Days in NH - mean (std)	663 (812)	813 (770)			26 (33)			<.001
1. < 6 months	32.3%	18.5%			99.2%			<.001
2. 6-12 months	13.3%	16.2%			0.8%			
3. 1-2 years	31.2%	38.5%			0%			
4. 3 years or more	23.2%	26.9%			0%			
Percent private room (%)	20.6%	23.6%			10.7%			.002

Item level applicability and response rates

Table 16 presents the number of respondents who skipped specific questions because they were not applicable (skipped) and the number of respondents who did not answer a question that was applicable to them (no response). (Complete distribution of responses to each of the questions in the two surveys are presented in Appendix M)

Twelve interview and eleven mail items had screeners to determine applicability. Hearing aid items and the item on being left sitting or laying too long in the same position had much lower applicability than other items; second highest were questions on dental and eye care – all items CMS requested. These low levels of applicability imply a need for substantially larger sample size. The required sample size to achieve a certain degree of inter-home discrimination also depends on the relative amount of within and between home variability (Appendix O). Item non-response ranged from 1-6% for the interview sample. It was higher for the mail sample with most items ranging from 1 – 11%. The religious needs item had the highest non-response in both the interview (6%) and mail (18%) sample. Skipped pages contributed to these higher rates in the mail survey; 5 people skipped 1 page and 14 individuals skipped 2 pages.

Table 16a. Percent of Residents Skipping and Not Responding to Survey Questions*

Variable names and labels in the combined data set (analysis data set)	Interview Sample N=439		Mail Sample n=127	
	Skipped	No response	Skipped	No response
C1. Rate food		0.2%		0.8%
C2. Ever eat in dining room		0.2%		0%
C3. Dining room meals	36.5%	0.7%	55.5%	0.8%
C4. Rate comfort of temperature		1.8%		1.6%
C5. Rate cleanliness of nursing home		1.4%		0.8%
C6. Rate safety, security		0.5%		1.6%
C7. Ever take pain medication		0.7%		7.0%
C8. Rate how well medicine helps with pain	23.2%	1.6%	12.5%	7.8%
C9. Rate how well staff helps with pain	23.2%	2.7%	12.5%	8.6%
C10. Rate speed of staff response to call for help		4.6%		8.6%
C11. Staff help you get dresses, take shower, or go to toilet		0%		7.0%
C12. Rate how gentle staff are when helping you	17.3%	0.5%	12.5%	7.0%
C13. Rate how respectful staff are to you		0.9%		7.8%
C14. Rate how well staff listen to you		2.3%		8.6%
C15. Rate how clearly staff explain things about your care		5.5%		9.4%
C16. Overall rating of care from staff		2.3%		8.6%
C17. Seen any doctor for medical care while in nursing home		2.3%		9.4%
C18. Rate medical care from doctors	16.4%	4.3%	20.3%	10.9%
C21. Is religion a part of your life?		1.8%		5.5%
C22. Rate staff help in meeting religious needs	24.8%	5.9%	28.1%	18.0%
C23. Overall rating of nursing home		1.8%		7.0%
C24. Area around room quiet at night		1.1%		6.3%
C25. Bothered by noise during the day		0.9%		3.9%
C27. Place for visiting in private	0%	3.4%	2.3%	6.3%
C28. Room set up so you can get things without help		3.0%		5.5%
C29. Can you turn over in bed without help?		0.9%		6.3%
C30. Ever left sitting or laying in same position so long it hurt	77.9%	1.4%	74.2%	6.3%

*If the "Skipped" cell is blank, the question did not have a preceding screener questions.

Table 16a. Percent of Residents Skipping and Not Responding to Survey Questions (continued)*

Variable names and labels in the combined data set (analysis data set)	Interview Sample N=439		Mail Sample n=127	
	Skipped	No response	Skipped	No response
C31. Did staff help you dress, take a shower, or bathe?		1.6%		3.9%
C32. Did staff make sure you had enough privacy while dressing, showering, bathing?	13.7%	1.8%	14.8%	7.0%
C33. Could you choose what time you went to bed?		2.3%		4.7%
C34. Could you choose what clothes you wore?		2.1%		4.7%
C35. Could you choose what activities you did there?		3.4%		10.2%
C36. Any volunteers or staff talk about kind of activities you like to do		2.1%		8.6%
C37. Enough organized activities on weekends		5.7%		21.1%
C38. Enough organized activities during week		3.4%		19.5%
C39. Recommend nursing home to others?		2.7%		5.5%
C40. How often feel worried		2.5%		5.5%
C41. How often feel happy		3.2%		7.0%
C42. Rate overall health		3.0%		4.7%
C43. Have roommate		4.3%		3.9%
C44. Rate your life when in nursing home		5.5%		10.2%
Year of Birth		9.6%		53.1%
Gender		0%		3.9%
Education		3.4%		7.0%
Ethnicity- Hispanic or Latino		3.0%		8.6%
Race		1.4%		3.9%

*If the “Skipped” cell is blank, the question did not have a preceding screener questions.

Table 16b. Percent of Residents Skipping and Not Responding to Survey Questions that were either only in the Interview or only the in Mail Survey*

Variable names and labels in the combined data set (analysis data set)	Interview Sample N=439		Mail Sample n=127	
	Skipped	No response	Skipped	No response
Items in the Interview Sample Only				
V1. Vignette 1 rating		7.3%		
V2. Vignette 2 rating		5.5%		
V3. Vignette 3 rating		7.7%		
I_q19. Using a hearing aid		0.7%		
I_q20. Any help from staff to keep hearing aid working	83.6%	1.4%		
I_q21. Rate staff help in keeping hearing aid working	88.6%	2.1%		
I_q22. Any help from staff with eye care		0.5%		
I_q23. Rate staff help with eye care	46.9%	1.1%		
I_q24. Any help from staff with dental care		0.7%		
I_q25. Rate staff help with dental care	55.8%	1.6%		
Items in Mail Sample Only				
M_q19. Any special therapy while in nursing home				4.7%
M_q20. Rate special therapy			2.3%	4.7%
M_q26. Have any visitors				3.9%
M_q49. Someone help you complete survey				4.7%
M_Q50A. Help - read the questions			72.7%	4.7%
M_Q50B. Help - write down answers I gave			72.7%	4.7%
M_Q50C. Help - Answered for me			72.7%	4.7%
M_Q50D. Help - Translated the questions			72.7%	4.7%
M_Q50E. Help – In some other way			72.7%	4.7%

*If the "Skipped" cell is blank, the question did not have a preceding screener questions.

Conclusion:

Several questions in both the interview and mailed survey were applicable to a relatively small proportion of the eligible respondents. These items should be considered for deletion if the survey needs to be shortened.

Cognitive impairment and item responses

Cognitive impairment can limit ability to respond to questions. To examine the extent to which the different measures of cognitive functioning predicted completeness of response patterns, we first calculated an index of response completeness. We did this by calculating the number of skipped items (Appropriately skipped items were not counted as non-responses. For example, if the respondent did not rate how gentle the staff were because the respondent did not get such help, that was not treated as a missing response.). We then examined the association between this index and the different measures of cognitive functioning. We found that 59% of interviewees answered all interview questions and 8% skipped four or more questions (Table 17). Item non-response was higher in the mail survey, perhaps due to inadvertently skipped pages. Only 25% of respondents answered every question in the mail survey and 26% skipped 4 or more items. The correlations between non-response and measures of cognition (Table 18) were statistically significant but weak, ranging from .11 to .29. For mail responders, the correlation between the number of skipped items and cognition as measured by the CPS was not significant ($r = -0.12$). Skipped pages in the mailed survey may account for this lack of relationship.

At the end of each interview, interviewers recorded their opinions on whether the respondent understood the questions. Response options were 1=yes always, 2=yes sometimes, and 3=no. When multiple visits were made to complete the interview the rating was done on each occasion and the average score across visits was used. These ratings were not correlated with non-response. The Short Blessed scores had a stronger independent association with the number of unanswered questions than other measures of cognition (CPS, vignette score, and interviewer observation [Table 19]). We were not able to assess the extent to which this stronger association with the Short Blessed was due to its placement. That is, it was administered at the end of the survey and this may have affected the extent to which it was related to non-response patterns.

We also assessed the association between cognitive functioning and the average of the 0 to 10 ratings across all rating questions in the survey provided by each respondent (Table 20). The correlations with the CPS were statistically significant but weak for interview sample (-.12, p=.01) and not significant for mail sample. The negative correlation indicates that more cognitively impaired persons may use lower mean ratings. No associations were found between the Short Blessed and mean ratings or between the vignette score and the mean ratings.

We also calculated the range of scores (on the 0 to 10 scale) used by each respondent across all rating items in the survey and found that that more cognitively impaired persons did not differ from less impaired residents in the range of values used.

Table 17. Frequency Distribution of Number of Questions Not Answered, by Survey Type

In-person Interviews

Number of applicable questions not answered	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	261	59.45	261	59.45
1	84	19.13	345	78.59
2	37	8.43	382	87.02
3	23	5.24	405	92.26
4	12	2.73	417	94.99
5	6	1.37	423	96.36
6	4	0.91	427	97.27
7	3	0.68	430	97.95
8	2	0.46	432	98.41
9	1	0.23	433	98.63
14	1	0.23	434	98.86
17	1	0.23	435	99.09
19	1	0.23	436	99.32
29	1	0.23	437	99.54
38	1	0.23	438	99.77
41	1	0.23	439	100.00

Mail Surveys

Number of applicable questions not answered	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	32	25.20	32	25.20
1	40	31.50	72	56.69
2	12	9.45	84	66.14
3	10	7.87	94	74.02
4	3	2.36	97	76.38
5	5	3.94	102	80.31
6	3	2.36	105	82.68
8	1	0.79	106	83.46
9	1	0.79	107	84.25
12	3	2.36	110	86.61
13	5	3.94	115	90.55
14	2	1.57	117	92.13
15	1	0.79	118	92.91
18	1	0.79	119	93.70
19	1	0.79	120	94.49
20	4	3.15	124	97.64
21	2	1.57	126	99.21
22	1	0.79	127	100.00

Table 18. Pearson correlations between number of missing responses and CPS score, Short Blessed Score and interviewer score

Interview Respondents

	Correlation Coefficient (sig.)
CPS score	0.11 (0.02)
Short Blessed score	0.29 (<.0001)
Vignette score	-0.20 (<.0001)
Interviewer score	0.05 (0.32)

Mail Respondents (only CPS score available)

	Correlation Coefficient (sig.)
CPS score	-0.12 (0.205)

Table 19. Regression coefficients between measures CPS score, Short Blessed Score, vignette score, and interviewer score and number of missing responses (dependent variable)

Variable	Parameter Estimate	Standard Error	Significance
CPS score	0.02	0.07	0.73
Short Blessed score	0.04	0.01	<.0001
Vignette score	-0.10	0.07	0.16
Interviewer score	0.06	0.13	0.64

Table 20a. Summary Statistics for Mean Rating Used by a Respondent Across All f 0-10 Rating Items Within the Survey

Sample	N	Mean	Std. Dev.	Min	Median	Max
Interview	439	7.98	1.48	2.43	8.29	10.00
Mail	127	8.00	1.47	3.67	8.22	10.00

Table 20b. Correlations between Measures of Cognitive Functioning and Mean 0-10 Rating Used by a Respondent Across All Ratings Items Within the Survey

	Interview Sample	Mail Sample
CPS	-0.12 (0.01)	-0.18 (0.06)
Short Blessed score	-0.07 (0.13)	
Vignette score	-0.03 (0.60)	

Comparison of Response Patterns between Samples

The interview sample consisted of long stay nursing home residents and the mail survey was sent to individuals who had recently completed a short post acute care nursing home stay. Table 21 compares the distribution of average scores across homes in the two samples. The distributions are similar in the two samples but the correlations between the two groups within a nursing home is quite low for many of the items, including most of the 0-10 ratings, suggesting that the experiences of one group are not a

good predictor the experiences of the other group. This could be because they have different needs and experiences, their expectations differ, and/or because of differences in characteristics described in Table 15 that might be related to reporting differences. Such differences should be evaluated in larger samples and when casemix models have been developed, those could be used to assess the extent to which the differences are due to differences between short and long stay patients. Until more data are available on this issue, caution should be used when combining information from these two groups of patients.

Table 21. Distribution of nursing home average item responses. N=12 (13 homes in interview sample, 12 homes in mail sample)

Label	Means of NH - Interview				Means of NH - Mail returns				Correlations	
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	R	p value
C1.Rate food	7.04	0.60	5.60	7.69	6.60	1.93	3.00	10.00	-0.339	0.2804
C3.Dining room meals - rate enjoyment	7.52	0.52	6.67	8.44	6.79	2.22	2.50	10.00	0.597	0.0895
C4.Rate comfort of temperature	7.40	0.53	6.27	8.36	7.71	1.33	5.67	10.00	-0.440	0.1525
C5.Rate cleanliness of nursing home	8.55	0.38	7.96	9.22	7.97	1.31	5.00	10.00	0.348	0.2670
C6.Rate safety, security	8.56	0.52	7.73	9.18	8.60	1.27	5.75	10.00	0.179	0.5774
C8.Rate how well medicine helps with pain	7.97	0.51	6.73	8.58	7.95	1.41	5.33	10.00	-0.477	0.1171
C9.Rate how well staff helps with pain	8.23	0.64	7.30	9.21	8.33	1.14	6.00	10.00	0.003	0.9937
C10.Rate speed of staff response to call for help	7.29	0.70	5.85	8.24	7.36	1.39	4.67	10.00	-0.267	0.4007
C12.Rate how gentle staff are when helping you	8.32	0.42	7.64	9.00	8.36	1.37	5.00	10.00	-0.093	0.7738
C13.Rate how respectful staff are to you	8.39	0.47	7.43	8.96	8.94	0.79	7.33	10.00	0.032	0.9220
C14.Rate how well staff listen to you	7.94	0.73	6.53	9.04	8.50	0.96	7.10	10.00	-0.341	0.2779
C15.Rate how clearly staff explain things about your care	7.88	0.47	7.08	8.54	8.14	1.13	6.33	10.00	-0.122	0.7058
C18.Rate medical care from doctors	8.34	0.63	6.92	9.20	7.68	1.66	3.67	10.00	-0.200	0.5322
C22.Rate staff help in meeting religious needs	8.20	0.71	6.94	9.24	6.90	1.85	2.67	9.40	-0.052	0.8730
C24.Area around room quiet at night	1.35	0.17	1.06	1.59	1.58	0.33	1.00	2.20	-0.084	0.7946
C25.Bothered by noise during the day	2.64	0.19	2.32	2.88	2.71	0.20	2.25	3.00	0.619	0.0318
C27.Place for visiting in private	1.23	0.14	1.00	1.50	1.31	0.38	1.00	2.33	0.475	0.1185
C28.Room set up so you can get things without help	1.67	0.16	1.38	1.92	1.43	0.24	1.00	1.80	0.007	0.9819
C30.Ever left sitting or laying in same position so long it hurts	2.11	0.28	1.67	2.60	2.29	0.69	1.00	3.00	-0.331	0.4234
C32.Do staff make sure you have enough privacy while dressing, showering, bathing?	1.09	0.08	1.00	1.24	1.16	0.24	1.00	1.75	-0.294	0.3535
C33.Can you choose what time you go to bed?	1.27	0.15	1.00	1.55	1.24	0.22	1.00	1.67	-0.300	0.3436
C34.Can you choose what clothes you wear?	1.16	0.10	1.05	1.44	1.08	0.14	1.00	1.50	0.050	0.8784
C35.Can you choose what activities you do here?	1.18	0.12	1.00	1.40	1.30	0.48	1.00	2.67	0.265	0.4048
C36.Any volunteers or staff talk about kind of activities you like to do	1.70	0.24	1.32	2.07	1.62	0.39	1.00	2.33	0.022	0.9458
C37.Enough organized activities on weekends	1.70	0.29	1.38	2.18	1.74	0.53	1.00	3.00	0.018	0.9559
C38.Enough organized activities during week	1.32	0.19	1.06	1.85	1.33	0.36	1.00	2.33	0.100	0.7570

Ceiling Effects

Table 21 presents the percent of respondents using the highest rating category for each item. As expected, we find less clustering in the highest response category on the 0-10 scale than for the other response formats. For the interviewed sample using the 0-10 rating scale, in 7 out of 18 items more than 40 % of respondents used the highest rating. The same was true for only 3 of the 18 items in the mail respondent sample. Ceiling effects were more pronounced among items with only 3 response options (Yes, Sometimes, No). This was most pronounced in the autonomy items (choose bedtime, choose clothes, choose activities) and the personal privacy item. The complete frequency distributions for all variables are included in Appendix M. Aside from the screening items, the items with the most skewed distributions included questions C31 to C35 and C38.

Table 21. Analysis of Ceiling Effects

	Interview Sample		Mail Sample	
Item	**% Best**	**% Yes**	**% Best**	**% Yes**
C1. Rating food	20.8		16.5	
C2. Ever eat in dining room		63.5		44.5
C3. Dining room meals	25.0		23.2	
C4. Rate comfort of temperature	23.2		29.4	
C5. Rate cleanliness of nursing home	41.6		33.1	
C6. Rate safety, security	47.6		0.0	
C7. Ever take pain medication		76.6		86.6
C8. Rate how well medicine helps with pain	30.3		33.3	
C9. Rate how well staff helps with pain	35.7		41.6	
C10. Rate speed of staff response to call for help	23.4		23.1	
C11. Staff help you get dressed, take shower, or go to toilet		82.7		86.6
C12. Rate how gentle staff are when helping you	39.9		43.7	
C13. Rate how respectful staff are to you	41.4		53.4	
C14. Rate how well staff listen to you	32.2		37.6	
C15. Rate how clearly staff explain things about your care	32.8		37.9	
C16. Overall rating of care from staff	42.4		39.3	
C17. Seen any doctor for medical care while in nursing home		83.2		77.6
C18. Rate medical care from doctors	40.8		33.0	
I_Q19. Using a hearing aid		15.8		-
I_Q20. Any help from staff to keep hearing aid working		66.7		-
I_Q21. Rate staff help in keeping hearing aid working	41.5		-	
I_Q22. Any help from staff with eye care		52.9		-
I_Q23. Rate staff help with eye care	43.4		-	
I_Q24. Any help from staff with dental care		43.8		-
I_Q25. Rate staff help with dental care	32.6		-	
M_Q19. Any special therapy while in nursing home		-		97.5
M_Q20. Rate special therapy		-		43.7
C21. Is religion a part of your life?		74.7		70.3
C22. Rate staff help in meeting religious needs	48.0		26.1	
C23. Overall rating of nursing home	41.8		38.7	
C24. Area around room quiet at night		76.5		58.3
C25. Bothered by noise during the day		11.0		4.1
M_Q26. Have any visitors		-		97.6
C27. Place for visiting in private		85.1		81.2
C28. Room set up so you can get things without help		71.4		69.4
C29. Can you turn over in bed without help?		78.6		79.2

	Interview Sample		Mail Sample	
Item	**% Best**	**% Yes**	**% Best**	**% Yes**
C30. Ever left sitting or laying in same position so long it hurt		33.0		24.0
C31. Did staff help you dress, take a shower, or bathe?		86.1		84.6
C32. Did staff make sure you had enough privacy while dressing, showering, bathing?		93.5		91.0
C33. Could you choose what time you went to bed?		83.5		84.4
C34. Could you choose what clothes you wore?		89.8		94.3
C35. Could you choose what activities you did there?		88.9		82.6
C36. Any volunteers or staff talk about kind of activities you like to do		58.8		64.1
C37. Enough organized activities on weekends		58.9		57.4
C38. Enough organized activities during week		80.7		73.8
C39. Recommend nursing home to others?*		58.1		59.5
C44. Rate your life when in nursing home	21.2		21.7	

*Rated definitely yes on a 4 point scale

Correlations with Rating Items

Table 22 presents the correlations between the overall rating items and other questions. Separate correlation matrices by sample (mail and interview) are presented in Appendix N.

Table 22. Pearson Correlations Matrix between Rating items and all survey items

c16: Overall Rating of staff

c23: Overall Rating of nursing home

c39: Would recommend NH to others

c40: How often worried

c41: How often feel happy

c42: Overall rating of health excellent (1) to poor (5)

c43: Percent have roommate 1=yes 2=no

c44: Rating of life (0-10)

Note 1. For all yes-no items, 1=yes, 2=no.

	c16	**c23**	**C39**	**c40**	**c41**
c1	0.3587	0.4392	0.3490	0.0874	-0.1528
Rate food	<.0001	<.0001	<.0001	0.0410	0.0004
c2	0.0159	-0.051	-0.067	0.0285	-0.0045
Ever eat in dining room	0.7114	0.2346	0.1199	0.5055	0.9173

	c16	c23	C39	c40	c41
c3	0.3912	0.4279	0.3321	0.1510	-0.1265
Dining room meals - rate enjoyment	<.0001	<.0001	<.0001	0.0067	0.0241
c4	0.4012	0.3877	0.2790	0.0655	-0.1369
Rate comfort of temperature	<.0001	<.0001	<.0001	0.1289	0.0015
c5	0.4648	0.5567	0.3812	0.1299	-0.1580
Rate cleanliness of nursing home	<.0001	<.0001	<.0001	0.0025	0.0002
c6	0.4874	0.4812	0.3853	0.1410	-0.1502
Rate safety, security	<.0001	<.0001	<.0001	0.0010	0.0005
c7	0.0328	0.0548	0.0144	0.2382	0.0792
Ever take pain medication	0.4458	0.2043	0.7399	<.0001	0.0676
c8	0.4347	0.3547	0.1958	0.1160	-0.1643
Rate how well medicine helps with pain	<.0001	<.0001	<.0001	0.0177	0.0008
c9	0.6375	0.5081	0.4144	0.1533	-0.2328
Rate how well staff helps with pain	<.0001	<.0001	<.0001	0.0018	<.0001
c10	0.6303	0.4578	0.3832	0.1743	-0.1614
Rate speed of staff response to call for help	<.0001	<.0001	<.0001	<.0001	0.0002
c11	0.0178	-0.018	-0.001	0.1561	-0.0622
Staff help you get dressed, take shower, or go to toilet	0.6784	0.6705	0.9907	0.0003	0.1503
c12	0.6155	0.5209	0.4172	0.1902	-0.1926
Rate how gentle staff are when helping you	<.0001	<.0001	<.0001	<.0001	<.0001
c13	0.6846	0.5263	0.4181	0.1380	-0.2140
Rate how respectful staff are to you	<.0001	<.0001	<.0001	0.0013	<.0001
c14	0.7427	0.5208	0.4166	0.1358	-0.2288
Rate how well staff listen to you	<.0001	<.0001	<.0001	0.0017	<.0001
c15	0.7294	0.4819	0.3924	0.1611	-0.2451
Rate how clearly staff explain things about your care	<.0001	<.0001	<.0001	0.0002	<.0001
c16	1.0000	0.5895	0.4205	0.1277	-0.2638
Overall rating of care from staff		<.0001	<.0001	0.0032	<.0001
c17	-0.025	0.0552	-0.032	0.1455	0.1157
Seen any doctor for medical care while in nursing home	0.5698	0.2035	0.4665	0.0008	0.0078
c18	0.4546	0.3711	0.2474	0.0346	-0.1403
Rate medical care from doctors	<.0001	<.0001	<.0001	0.4758	0.0039

	c16	c23	c39	c40	c41
c21	-0.150	-0.091	-0.158	0.0098	0.1532
Is religion a part of your life?	0.0005	0.0338	0.0002	0.8209	0.0004
c22	0.4393	0.4680	0.2645	-0.0246	-0.0917
Rate staff help in meeting religious needs	<.0001	<.0001	<.0001	0.6399	0.0818
c23	0.5895	1.0000	0.5564	0.1335	-0.2153
Overall rating of nursing home	<.0001		<.0001	0.0020	<.0001
c24	-0.209	-0.210	-0.161	-0.093	0.0208
Area around room quiet at night	<.0001	<.0001	0.0002	0.0301	0.6309
c25	0.2039	0.2205	0.1909	0.1346	-0.1541
Bothered by noise during the day	<.0001	<.0001	<.0001	0.0017	0.0003
c27	-0.246	-0.211	-0.171	-0.069	0.1251
Place for visiting in private	<.0001	<.0001	<.0001	0.1166	0.0042
c28	-0.163	-0.279	-0.141	-0.126	0.1306
Room set up so you can get things without help	0.0002	<.0001	0.0011	0.0035	0.0026
c29	-0.083	-0.120	-0.104	-0.194	0.0827
Can you turn over in bed without help?	0.0556	0.0050	0.0158	<.0001	0.0560
c30	0.1423	0.1898	0.3804	0.1906	-0.1630
Ever left sitting or laying in same position so long it hurts	0.1311	0.0431	<.0001	0.0431	0.0844
c31	0.0552	-0.005	-0.029	0.1557	-0.1624
Do staff help you dress, take a shower, or bathe?	0.2016	0.9170	0.5077	0.0003	0.0002
c32	-0.274	-0.233	-0.226	-0.084	0.2208
Do staff make sure you have enough privacy while dressing, showering, bathing ?	<.0001	<.0001	<.0001	0.0726	<.0001
c33	-0.106	-0.163	-0.136	-0.085	0.0774
Can you choose what time you go to bed?	0.0144	0.0001	0.0016	0.0482	0.0743
c34	-0.162	-0.158	-0.155	0.0094	0.1016
Can you choose what clothes you wear?	0.0002	0.0002	0.0003	0.8285	0.0188
c35	-0.160	-0.222	-0.147	-0.022	0.0909
Can you choose what activities you do here?	0.0003	<.0001	0.0007	0.6086	0.0379

	c16	c23	c39	c40	c41
c36	-0.175	-0.145	-0.133	0.0425	0.1277
Any volunteers or staff talk about kind of activities you like to do	<.0001	0.0008	0.0020	0.3271	0.0032
c37	-0.193	-0.284	-0.283	-0.169	0.1720
Enough organized activities on weekends	<.0001	<.0001	<.0001	0.0001	0.0001
c38	-0.280	-0.308	-0.340	-0.126	0.1990
Enough organized activities during week	<.0001	<.0001	<.0001	0.0041	<.0001
c39	0.4205	0.5564	1.0000	0.1878	-0.1934
Recommend nursing home to others?	<.0001	<.0001		<.0001	<.0001
c40	0.1277	0.1335	0.1878	1.0000	-0.1927
How often feel worried	0.0032	0.0020	<.0001		<.0001
c41	-0.264	-0.215	-0.193	-0.193	1.0000
How often feel happy	<.0001	<.0001	<.0001	<.0001	
c42	-0.147	-0.140	-0.167	-0.236	0.2405
Rate overall health	0.0007	0.0012	<.0001	<.0001	<.0001
c43	-0.102	-0.160	-0.094	-0.041	0.0433
Currently have roommate	0.0192	0.0002	0.0306	0.3416	0.3195
c44	0.2908	0.3442	0.3063	0.2861	-0.3846
Rate your life now	<.0001	<.0001	<.0001	<.0001	<.0001

Factor Analyses

We first conducted exploratory factor analyses of all the questions, except for the global ratings, separately for mail and interview respondents. Those analyses resulted in very discrepant patterns, with multiple factors in each sub-sample. Although the data suggest that the experiences of long and short stay residents are different and they potentially have different patterns of correlations, the sample of mail (short stay) residents is too small to develop precise estimates of factor analysis coefficients. Thus, we present results from only the combined sample. We used a principal factor method and used R squared as the initial communality estimates. We then estimated an oblique (PROMAX; power = 3) rotation. These analyses yielded two factors with eigenvalues greater than 1.0. Those two factors explained 77% of the variability in responses. When rotated, those two factors had a correlation of -0.47. The eigenvalues for the first 5 factors are presented in Table 23.

Table 23. Eigenvalues

	Eigenvalue	Difference	Proportion of variance explained	Cumulative variance explained
1	8.77878460	7.60594402	0.6801	0.6801
2	1.17284058	0.27520812	0.0909	0.7710
3	0.89763246	0.10648909	0.0695	0.8405
4	0.79114336	0.15682458	0.0613	0.9018
5	0.63431878	0.14372086	0.0491	0.9510

Table 24 presents the factor pattern coefficients for the two factor solution. The results in Table 24 indicate that the responses clustered generally in ways that were anticipated, but there were no strong empirical grouping of items that were substantively consistent. Analyses (not shown) of the interview only sample yielded similar results. In that subsample of respondents, there were also two eigenvalues greater than 1.0 (7.0 and 1.2) that explained 75% of the variance. There were some differences in the patterning of the factor pattern coefficients, but the results were generally similar.

Table 24. Rotated Factor Pattern (Standardized Regression Coefficients)

			Factor1	Factor2
c1	Rate food	Life	0.34968	-0.24291
c3	Dining room meals - rate enjoyment	Life	0.38702	-0.32939
c4	Rate comfort of temperature		**0.55255**	-0.02371
c5	Rate cleanliness of nursing home		**0.65581**	-0.04126
c6	Rate safety, security	Life	**0.61790**	-0.05137
c8	Rate how well medicine helps with pain		**0.60437**	0.06724
c9	Rate how well staff helps with pain		**0.76119**	-0.00010
c10	Rate speed of staff response to call for help		**0.68703**	-0.05177
c12	Rate how gentle staff are when helping you		**0.80088**	0.05202
c13	Rate how respectful staff are to you		**0.83744**	0.09629
c14	Rate how well staff listen to you		**0.83912**	0.06041
c15	Rate how clearly staff explain things about your care		**0.77273**	0.00526
c16	Overall rating of care from staff	Global	**0.8300**2	0.02468
c18	Rate medical care from doctors		**0.41613**	-0.18267
c22	Rate staff help in meeting religious needs	Life	0.33830	-0.32576
c23	Overall rating of nursing home	Global	**0.58402**	-0.25331
c24	Area around room quiet at night		-**0.22256**	0.05484
c25	Bothered by noise during the day		0.18178	-0.16643
c27	Place for visiting in private	Life	-0.21023	0.21689
c28	Room set up so you can get things without help		-0.20741	0.07664
c30	Ever left sitting or laying in same position so long it hurts		**0.31122**	-0.01791
c32	Do staff make sure you have enough privacy…	Life	**-0.32498**	0.05286
c33	Can you choose what time you go to bed?	Life	-0.08747	**0.23556**
c34	Can you choose what clothes you wear?	Life	-0.03122	**0.36662**
c35	Can you choose what activities you do here?	Life	0.01657	**0.48833**
c36	Any volunteers or staff talk about kind of activities you like to do	Life	-0.00541	**0.29647**
c37	Enough organized activities on weekends	Life	-0.00263	**0.56349**
c38	Enough organized activities during week	Life	0.00444	**0.63745**
c39	Recommend nursing home to others?	Global	0.40443	-0.30469
c44	Rate your life now		0.29712	-0.18767

Factor Structure (Correlations)

		Factor1	Factor2
c1	Rate food	0.46312	-0.40621
c3	Dining room meals - rate enjoyment	0.54084	-0.51012
c4	Rate comfort of temperature	0.56362	-0.28174
c5	Rate cleanliness of nursing home	0.67508	-0.34751
c6	Rate safety, security	0.64189	-0.33992
c8	Rate how well medicine helps with pain	0.57297	-0.21499
c9	Rate how well staff helps with pain	0.76124	-0.35556
c10	Rate speed of staff response to call for help	0.71120	-0.37260
c12	Rate how gentle staff are when helping you	0.77659	-0.32197
c13	Rate how respectful staff are to you	0.79247	-0.29478
c14	Rate how well staff listen to you	0.81091	-0.33145
c15	Rate how clearly staff explain things about your care	0.77028	-0.35560
c16	Overall rating of care from staff	0.81850	-0.36293
c18	Rate medical care from doctors	0.50143	-0.37700
c22	Rate staff help in meeting religious needs	0.49042	-0.48374
c23	Overall rating of nursing home	0.70231	-0.52603
c24	Area around room quiet at night	-0.24817	0.15877
c25	Bothered by noise during the day	0.25950	-0.25132
c27	Place for visiting in private	-0.31151	0.31506
c28	Room set up so you can get things without help	-0.24320	0.17349
c30	Ever left sitting or laying in same position so long it hurts	0.31959	-0.16325
c32	Do staff make sure you have enough privacy while dressing, showering, bathing?	-0.34967	0.20463
c33	Can you choose what time you go to bed?	-0.19747	0.27641
c34	Can you choose what clothes you wear?	-0.20243	0.38120
c35	Can you choose what activities you do here?	-0.21147	0.48059
c36	Any volunteers or staff talk about kind of activities you like to do	-0.14386	0.29899
c37	Enough organized activities on weekends	-0.26577	0.56472
c38	Enough organized activities during week	-0.29324	0.63538
c39	Recommend nursing home to others?	0.54671	-0.49355
c44	Rate your life now	0.38476	-0.32642

Given the lack of clear empirical clustering of items, we developed and tested several sets of scales, based primarily on the substantive content of the items. Because several items could be interpreted as reflecting more than one aspect of care, we tested multiple combinations. Those item combinations and the internal consistency of the items in the scale (coefficient Alpha) are presented in Table 25.

Table 25. Internal consistency (coefficient Alpha-standardized) for different scale configurations

Scale	Items in the scale	Sample Used	Alpha
Environment - original	C1,C3,C4,C5,C6,C24R,C25,C27R	Pooled	0.72
		Interview	0.72
		Mail	0.80
Environment - delete c27	C1,C3,C4,C5,C6,C24R,C25	Pooled	0.74
		Interview	0.72
		Mail	0.80
Environment - delete food items c1&c3	C4,C5,C6,C24R,C25,C27R	Pooled	0.64
		Interview	0.61
		Mail	0.72
***Environment - add c28**	**C1,C3,C4,C5,C6,C24R,C25,C27R,C28R**	**Pooled**	**0.73**
		Interview	**0.71**
		Mail	**0.80**
Environment - food only	C1,C3	Pooled	0.75
		Interview	0.72
		Mail	0.81
Environment - privacy	C27R, C32R	Pooled	0.36
		Interview	0.41
		Mail	0.13
Care - Original	C8,C9,C10,C12,C30	Pooled	0.76
		Interview	0.77
		Mail	0.74
Care - delete c30	C8,C9,C10,C12	Pooled	0.81
		Interview	0.81
		Mail	0.79
Care - add c32	C8,C9,C10,C12,C30,C32R	Pooled	0.75
		Interview	0.75
		Mail	0.75
Care - add c6	C8,C9,C10,C12,C30,C6	Pooled	0.80
		Interview	0.80
		Mail	0.81
Care - add c36	C8,C9,C10,C12,C30,C36R	Pooled	0.71
		Interview	0.71
		Mail	0.71
***Care - delete c30 & add c32**	**C8,C9,C10,C12,C32R**	**Pooled**	**0.77**
		Interview	**0.79**
		Mail	**0.73**
Care -delete c30 & add c36	C8,C9,C10,C12,C36R	Pooled	0.73
		Interview	0.74
		Mail	0.73
***Communication/Respect - original**	**C13,C14,C15**	**Pooled**	**0.86**
		Interview	**0.86**
		Mail	**0.89**
Communication/Respect - add c36	C13,C14,C15,C36R	Pooled	0.74
		Interview	0.73
		Mail	0.79
Communication/Respect - add c10, c12	C13,C14,C15,C10,C12	Pooled	0.89
		Interview	0.88
		Mail	0.90

*Recommended scale for reporting

Table 25. Internal consistency (coefficient Alpha-standardized) for different scale configurations (continued)

Scales	Items in the scale	Sample Used	Alpha
Autonomy - original	C28,C32,C33,C34,C35	Pooled	0.52
		Interview	0.50
		Mail	0.45
Autonomy - delete c32	C28,C33,C34,C35	Pooled	0.49
		Interview	0.51
		Mail	0.44
***Autonomy - Short: Choice**	**C33,C34,C35**	**Pooled**	**0.56**
		Interview	**0.60**
		Mail	**0.40**
Activities - original	C36,C37,C38	Pooled	0.57
		Interview	0.53
		Mail	0.75
***Activities - delete c36**	**C37,C38**	**Pooled**	**0.64**
		Interview	**0.60**
		Mail	**0.81**

***Recommended scale for reporting**

The study team reviewed the different possible item combinations and considered the estimated reliability in the mail, interview, and combined samples as well as the face validity of each item for the scale considered. The scales selected as recommended reporting scales are indicated with "*" and bolded in Table 25. The team also thought that C1 and C3, questions about food, could be treated as a separate subscale if there were interest in tracking that aspect of nursing home quality separately.

Once the team made decisions regarding the reporting scales to be used, we assessed several characteristics of these scales. Table 26 presents descriptive statistics for each of the specified scales.

Table 26. Descriptive statistics and Number of missing items for the scales

Table 26a. All subjects (pooled interview and mail sample) N=566

Scales	Num of items per scale	Summary Statistics for Scales		Number of Missing Items for each scale						
		Mean	SD	Mean	Frequency Distribution (Percent)					
					0	**1**	**2**	**3**	**4**	**5**
Environment	9	5.47	1.01	0.16	510 (90.1%)	41 (7.2%)	5 (0.9%)	0	9 (1.6%)	1 (0.2%)
Care	5	6.90	1.66	0.17	504 (89.1%)	45 (8.0%)	8 (1.4%)	0	7 (1.2%)	2 (0.4%)
Communication and Respect	3	8.14	1.93	0.13	522 (92.2%)	28 (5.0%)	5 (0.9%)	11 (1.9%)		
Autonomy	3	2.80	0.40	0.10	528 (93.3%)	27 (4.77%)	1 (0.2%)	10 (1.8%)		
Activities	2	2.51	0.66	0.16	510 (90.1%)	22 (3.9%)	34 (6.0%)			

Table 26b. Interview sample, N=439

Scales	Num of items per scale	Summary Statistics for Scales		Number of Missing Items for each scale						
		Mean	SD	Mean	Frequency Distribution (Percent)					
					0	**1**	**2**	**3**	**4**	**5**
Environment	9	5.47	0.97	0.13	399 (91.0%)	32 (7.3%)	4 (0.9%)	0	3 (0.7%)	1 (0.2%)
Care	5	6.88	1.65	0.11	397 (90.4%)	35 (8.0%)	7 (1.6%)	0	0	0
Communication and Respect	3	8.06	1.99	0.09	408 (92.9%)	26 (5.9%)	3 (0.7%)	2 (0.5%)		
Autonomy	3	2.80	0.42	0.08	415 (94.5%)	19 (4.3%)	0	5 (1.1%)		
Activities	2	2.51	0.66	0.09	410 (93.4%)	18 (4.1%)	11 (2.5%)			

Table 26c. Mail sample, N=127

Scales	Num of items per scale	Summary Statistics for Scales		Number of Missing Items for each scale						
		Mean	SD	Mean	Frequency Distribution (Percent)					
					0	**1**	**2**	**3**	**4**	**5**
Environment	9	5.47	1.14	0.28	111 (87.4%)	9 (7.1%)	1 (0.8%)	0	6 (4.7%)	0
Care	5	6.97	1.79	0.39	107 (84.3%)	10 (7.9%)	1 (0.8%)	0	7 (5.5%)	2 (1.6%)
Communication and Respect	3	8.47	1.67	0.26	114 (90.0%)	2 (1.6%)	2 (1.6%)	9 (7.1%)		
Autonomy	3	2.82	0.34	0.20	113 (89.0%)	8 (6.3%)	1 (0.8%)	5 (3.9%)		
Activities	2	2.50	0.67	0.39	100 (78.7%)	4 (3.2%)	23 (18.1%)			

Tables 27a, b, and c present the present the correlations between the scales and the ratings. The most striking pattern in these data is the large correlation between the Communication and Respect scale and the overall rating of care from nursing home staff. The scale that generally had the lowest correlations with the ratings is the Autonomy scale. The Environment and Communication and Respect scales tended to have the strongest correlations with the overall rating of the nursing home.

Tables 27d, e, and f. present the correlations between the composites. The composite scores tend to be highly correlated, indicating that they provide information that is to a large degree shared across scales. That is, if the only goal of the survey data were to develop scores that discriminated between nursing homes, without regard to the substantive content of items, it probably would be possible to develop a smaller number of composites that provided comparable information about inter-home variability.

Table 27. Correlations between scales and overall ratings

Table 27a. All subjects (pooled interview and mail sample) N=566

Ratings	Scales				
	Environment	Care	Communication & Respect	Autonomy	Activities
Rating of Care from NH Staff	0.55 ***	0.65 ***	0.80 ***	0.20 ***	0.27 ***
Rating of Medical Care from MD	0.38 ***	0.44 ***	0.45 ***	0.20 ***	0.30 ***
Overall Rating of Nursing Home	0.62 ***	0.49 ***	0.57 ***	0.24 ***	0.34 ***
Would Recommend NH to other	0.47 ***	0.34 ***	0.45 ***	0.20 ***	0.36 ***
Overall Rating of Life at NH	0.32 ***	0.36 ***	0.33 ***	0.12 **	0.29 ***

Table276b. Interview sample, N=439

Ratings	Scales				
	Environment	Care	Communication & Respect	Autonomy	Activities
Rating of Care from NH Staff	0.55 ***	0.63 ***	0.79 ***	0.20 ***	0.21 ***
Rating of Medical Care from MD	0.39 ***	0.49 ***	0.49 ***	0.19 ***	0.27 ***
Overall Rating of Nursing Home	0.57 ***	0.47 ***	0.55 ***	0.24 ***	0.28 ***
Would Recommend NH to other	0.45 ***	0.33 ***	0.42 ***	0.20 ***	0.34 ***
Overall Rating of Life at NH	0.28 ***	0.30 ***	0.30 ***	0.12 *	0.28 ***

Table 27c. Mail sample, N=127

Ratings	Scales				
	Environment	Care	Communication & Respect	Autonomy	Activities
Rating of Care from NH Staff	0.60 ***	0.73 ***	0.87 ***	0.21 *	0.54 ***
Rating of Medical Care from MD	0.36 ***	0.35 ***	0.35 ***	0.26 *	0.42 ***
Overall Rating of Nursing Home	0.75 ***	0.56 ***	0.72 ***	0.27 **	0.56 ***
Would Recommend NH to other	0.57 ***	0.40 ***	0.61 ***	0.20 *	0.45 ***
Overall Rating of Life at NH	0.45 ***	0.58 ***	0.49 ***	0.15	0.36 ***

Note: * p <.05
** p <.01
*** p <.001

Table 27d. All subjects (pooled interview and mail sample) N=566

	Environment	Care	Communication & Respect	Autonomy	Activities
Environment	1.00				
Care	0.67	1.00			
Communication	0.61	0.75	1.00		
Autonomy	-0.36	-0.34	-0.39	1.00	
Activities	-0.26	-0.27	-0.27	0.31	1.00

Table 27e. Interview sample, N=439

	Environment	Care	Communication & Respect	Autonomy	Activities
Environment	1.00				
Care	0.69	1.00			
Communication	0.63	0.75	1.00		
Autonomy	-0.34	-0.32	-0.39		
Activities	-0.25	-0.24	-0.24	0.28	1.00

Table 27f.Mail sample, N=127

	Environment	Care	Communication & Respect	Autonomy	Activities
Environment	1.00				
Care	0.64	1.00			
Communication	0.56	0.77	1.00		
Autonomy	-0.43	-0.41	-0.42	1.00	
Activities	-0.31	-0.40	-0.39	0.40	1.00

Note: All correlation coefficients are significant at p<.001

Table 28 presents the home level reliability of the scales for all subjects and for the separate samples. These results indicate that with the sample sizes from the field test, most of the scales achieved only moderate reliability at the home level and that from 13 to 100 respondents would be needed for the scales to achieve a home level reliability of 0.70, depending on the sample, or mode of data collection. The F-tests for individual questions are presented in Appendix O. The nature of the study sample should be considered when interpreting these statistics. Since participation in the field test was voluntary, participating homes might be more homogeneous that a more random sample of homes in more diverse locations.

If larger samples from a more diverse sample of homes confirms that more respondents are needed than many homes have at a particular time, samples may need to be collected sequentially over longer periods of time to accumulate adequate samples for making inferences about individual homes.

When making comparisons among facilities that might have too few eligible cases for sufficiently precise estimates in a single year, the question often arises about whether one should replace the standard method of calculating statistical precision with that derived from the finite-population sampling model (FPSM). The main criticism of the FPSM method is that typically one is not only interested in evaluating the experiences of the particular group of residents in the institution during the assessment period. Rather,

their experiences are the basis for making inferences about the general performance of the institution, as a basis for evaluating general institutional quality, and predicting the outcomes for future residents. For this reason, inferences about relative performance are typically conducted under the standard statistical assumptions that permit inference to large or theoretically infinite universes.

Table 28. Nursing Home Level Reliability of Scales

Table 28a. For all subjects

Scales	F	Group Reliability	Total non-missing N	Average N per NH (k=13)	Num of subjects need to reach Reliability of 0.70
Environment	2.06	0.51	566	43.5	95.8
Care	2.59	0.61	558	42.9	63.0
Communication & Respect	2.36	0.58	555	42.7	73.2
Autonomy	1.72	0.42	556	42.8	138.6
Activities	2.85	0.65	532	40.9	51.6

Table 28b. For Interview Sample only

Scales	F	Group Reliability	Total non-missing N	Average N per NH (k=13)	Num of subjects need to reach Reliability of 0.70
Environment	1.85	0.46	439	33.8	92.7
Care	2.56	0.61	435	33.5	50.0
Communication & Respect	2.40	0.58	436	33.5	55.9
Autonomy	1.96	0.49	434	33.4	81.1
Activities	3.60	0.72	428	32.9	29.5

Table 28c. For Mail Sample Only

Scales	F	Group Reliability	Total non-missing N	Average N per NH (k=12)	Num of subjects need to reach Reliability of 0.70
Environment	2.81	0.64	127	10.6	13.6
Care	2.40	0.58	123	10.3	17.1
Communication & Respect	1.86	0.46	118	9.8	26.7
Autonomy	1.62	0.38	122	10.2	38.3
Activities	1.63	0.39	104	8.7	32.1

Notes: (1) Group-level reliability = (F-1)/F;
(2) Number of subjects need to reach Reliability of 0.70: Calculated with the Spearman-Brown prophecy formula using the average num of respondents/group.

Conclusions and Recommendations

The Nursing Home CAHPS team successfully developed, refined, and tested procedures for conducting an interview survey of current long term nursing home residents and a mail survey of discharged short term residents. Aside from the logistics of contacting the nursing homes, eliciting participation, drawing a sample, and arranging time to interview residents, the most challenging aspect of this project was collecting data from a group of individuals with a relatively high probability of cognitive impairment. Several instruments (CPS, Short Blessed, vignettes) for testing the types of cognitive skills necessary to complete an interview were tested, and, the Short Blessed, was particularly successful at identifying residents who were likely to have difficulty answering all the questions, even though they could complete the interview. However, interviewers felt that administering the Short Blessed scale was burdensome and intrusive. While earlier testing indicated that correctly scaled vignettes helped identify respondents with lower levels of item nonresponse , their performance was not as strong in the field test. In addition, interviewers reported that some residents had difficulty with the hypothetical nature of the vignette questions. Thus, we are not recommending their continued inclusion. The interviewers said that the strategy of stopping the interview whenever a respondent was unable to answer any three questions in a row was the least intrusive and stressful way to identify respondents who could not be interviewed. The team recommends that the CPS data in addition to this procedure be used in future interview studies. Other conclusions and recommendations regarding the implementation and costs of data collection activities are summarized in parts I and II of this report.

The instruments tested performed well. If they needed to be shortened, there were several questions that apply to a relatively small subset of respondents that could be dropped. In addition, there were several questions that the interviewers thought were interpreted differently than intended and/or that were difficult to answer, as described in Part I of this report. The team recommends that Question C30 (Ever left sitting or laying…) be dropped because it applied to so few residents and did not seem to fit well in any of the developed scales. Further work is needed to refine other questions, such as --C17 and C18 (getting care from doctor), C20 and C21 (help with hearing aid), C26 and C27 (religious support), C32 (how room was set-up), and C40 (asking about things to do). (These items and the issues that arose during the interviews are discussed in Section I of this report.) Of these items, only Q27 about religious needs is included in one of the recommended scales. However, the team thought that because of programmatic or constituent interest in the issues addressed by these questions, they should remain in the survey until alternative questions are developed and tested.

CODEBOOK FOR CAHPS NURSING HOME SURVEY: LONG STAY RESIDENT INSTRUMENT SPECIFICATIONS

Composite 1. Environment (8 items)

In this composite, five questions use the 0 to 10 rating scale in which "10" represents the highest quality and "0" represents the lowest quality, and three questions use a "yes", "sometimes," and "no" scale in which higher numbers represent better quality.

Q1: "What number would you use to rate the food here at the nursing home?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q3: "When you eat in the dining room, what number would you use to rate how much you enjoy mealtimes?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q4: "What number would you use to rate how comfortable the temperature is in the nursing home?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q5: "What number would you use to rate how clean the nursing home is?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q6: "What number would you use to describe how safe and secure you feel in the nursing home?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q18: "Is the area around your room quiet at night?"
- No = 1
- Sometimes = 2
- Yes = 3

Q19: "Are you bothered by noise in the nursing home during the day?" (*Note: reverse coding*)
- Yes = 1
- Sometimes = 2
- No = 3

Q20: "If you have a visitor, can you find a place to visit in private?"
- No = 1
- Sometimes = 2

- Yes = 3

Composite 2. Care (5 items)

In this composite, four questions use the 0 to 10 rating scale in which "10" represents the highest quality and "0" represents the lowest quality, and one question uses a "yes," "no," "sometimes," scale in which a higher number represents better quality.

Q.8: "What number would you use to rate how well the medicine worked to help with aches or pain?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q9: "What number would you use to rate how well the staff help you when you have pain?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q10: "What number would you use to rate how quickly the staff come when you call for help?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q12: "What number would you use to rate how gentle the staff are when they're helping you?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q29: "Do the staff make sure you have enough personal privacy when you dress, take a shower, or bathe?"
- No = 1
- Sometimes = 2
- Yes = 3

Composite 3. Communication/Respect (3 items)

In this composite, three questions use the 0 to 10 rating scale in which "10" represents the highest quality and "0" represents the lowest quality.

Q13: "What number would you use to rate how respectful the staff are to you?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q14: "What number would you use to rate how well the staff listen to you?"
- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q15: "What number would you use to rate how clearly the staff explain things about your care to you?"

- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Composite 4. Autonomy (3 items)

In this comp osite, three questions use the "yes," "no," "sometimes," scale in which a higher number represents better quality.

Q30: "Can you choose what time you go to bed?"
- No = 1
- Sometimes = 2
- Yes = 3

Q31: "Can you choose what clothes you wear?"
- No = 1
- Sometimes = 2
- Yes = 3

Q32: "Can you choose what activities you do here?"
- No = 1
- Sometimes = 2
- Yes = 3

Composite 5. Activities (2 items)

In this composite, two questions use the "yes," "no," "sometimes," scale in which a higher number represents better quality.

Q33: "Are there enough organized activities for you to do on the weekends?"
- No = 1
- Sometimes = 2
- Yes = 3

Q34: "Are there enough organized activities for you to do during the week?"
- No = 1
- Sometimes = 2
- Yes = 3

Global Ratings & Items

Two questions use the 0 to 10 rating scale in which "10" represents the highest quality and "0" represents the lowest quality, and one question uses a "definitely no," "probably no," "probably yes," and "definitely yes" scale in which a higher number represents better quality.

Q16: "Overall, what number would you use to rate the care you get from the staff?" (0-10)

- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q17: "Overall, what number would you use to rate this nursing home?"

- 10 = 10 (highest quality)
- 0 = 0 (lowest quality)

Q35: "Would you recommend this nursing home to others?"

- Definitely No = 1
- Probably No = 2
- Probably Yes = 3
- Definitely Yes = 4

AMERICAN INSTITUTES FOR RESEARCH®

CAHPS® Survey for Family Members of Nursing Home Residents

Final Report

July 2008

Prepared for:

Judith Sangl, ScD.
Project Officer
Agency for Healthcare Research and Quality
540 Gaither Road, 3rd Floor
Rockville, MD 20850

Prepared by:

Elizabeth Frentzel, M.P.H.
Chris Evensen, M.S.
San Keller, Ph.D.
Steven A. Garfinkel, Ph.D.
American Institutes for Research
101 Conner Drive, Suite 301
Chapel Hill, NC 27514

Executive Summary

The CAHPS II team, comprising the American Institutes of Research (AIR), the RAND Corporation, and Yale University/Harvard University/University of Massachusetts Center for Survey Research (CSR), with assistance from Westat, developed, tested, and fielded a nursing home survey for family members or responsible parties. The CAHPS Nursing Home *Family* Survey ("family survey") complements the CAHPS Nursing Home *Resident* Survey ("resident survey") in that it was not designed to provide the same information that the resident survey provides. The family survey does not ask respondents to report on the residents experience; rather, it asks respondents to report on their own experiences.

Survey Development Process

The development of this survey instrument involved several steps:

1. Publishing a call for measures in the Federal Register, including requesting survey, items or domains from other home health surveys
2. Reviewing previous literature and other nursing home surveys
3. Conducting and focus groups with family members of nursing home residents.
4. Developing a draft survey instrument
5. Cognitively testing the draft survey in English
6. Obtaining input from the nursing home technical expert panel (TEP) to review candidate items and composites
7. Refining the survey instrument
8. Pilot testing the survey with 15 nursing homes in Texas
9. Analyzing the pilot data and revising the survey
10. Obtaining input from the TEP to finalize the survey.

The composite structure was determined using psychometric statistical techniques, including factor analysis and multi-trait analysis. The case-mix adjustment analysis included multiple regression and variance component analyses.

Final Composite Structure

As part of the standard CAHPS survey development process, composite measures are developed from items measuring the same aspect of care. The CAHPS team conceptually defined

composites representing our hypothesis about the "structure" that the survey data would reflect. We then conducted a confirmatory factor analysis (CFA) to determine whether the pilot data were consistent with the hypothesized structure. The CFA did not wholly support the structure we had hypothesized. Therefore, we conducted an exploratory factor analysis to help us define reliable and valid composites and to help us identify items that should be revised or deleted. The composite structure that AIR proposed and the team agreed to presents the best combination of conceptual properties and statistical support. The proposed composite structure is listed in Table 1 (this is also in Table 7 on page 13 in full report).

Table 1. Final Composites and Items

Q#	Composite or Item Handle
	Meeting Basic Needs: Help with eating, drinking, toileting [a]
19	Wait too long for help with eating
21	Wait too long for help with drinking
23	Wait too long for help with toileting
	Nurses/Aides' Kindness/ Respect Towards Resident
12	Nurses/Aides treat resident with respect
13	Nurses/Aides treat resident with kindness
14	Nurses/Aides really cared about. resident
15	Nurses/Aides rude to resident
26	Nurses/Aides appropriate with violent resident
	Nursing Home Provides Info/ Encourages Respondent Involvement
29	Nurses/Aides give respondent information about resident
30	Nurses/Aides explain things to respondent
31	Nurses/Aides discourage respondents questions
43	Respondent stops self from complaining
45	Respondent involved in decisions about care
53	Respondent given Info. about payments/expenses
	Nursing Home Staffing, Care of Belongings, & Cleanliness
11	Can find a nurse or aide
51	Enough nurses/ aides?
32	Room looks/smells clean
24	Resident looks/ smells clean
35	Public areas look/smell clean
37	FM's medical belongings lost
39	FM's clothes lost

This composite structure excludes 10 substantive items that did not perform well. Table 2 lists the excluded items and explains why they were excluded.

Table 2. Excluded Items

Q#	Items Excluded from Composites	Reasons for Excluding
16	Nurses/Aides treat resident rough	Marginal measurement characteristics.
17	Another resident rude to respondent's family member	Marginal measurement characteristics.
27	Nurses/Aides treat respondent with respect	Not uniquely and strongly related to its composite. Other items tap concept.
33	Noise level around room	Not uniquely and strongly related to its composite.
34	Places to talk to resident in private	Not uniquely and strongly related to its composite. Other items tap concept.
36	Nurses/Aides didn't protect resident's modesty	Marginal measurement characteristics.
42	Staff handled respondent's concerns satisfactorily	Includes term "satisfaction" which is not consistent with CAHPS methods; also low NH-reliability
46	Nursing Home has care conferences	Care conferences are required in Canada ONLY.
48	Management handled complaints satisfactorily	Includes term, "satisfaction" which is not consistent with CAHPS methods; also low NH-reliability
58	Staff did not give respondent info. due to privacy laws	Marginal measurement properties.

Recommended Case-Mix Adjusters

After conducting numerous analyses, we concluded that using four variables should be used as case-mix adjusters for the family survey: respondent age, respondent education, whether resident was permanently in the nursing home, and whether the resident was capable of making decisions.

Contents

APPENDICES

The Development of a CAHPS® Survey for Family Members of Nursing Home Residents

Project Description

Over the last decade, the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) program, has led the way in developing a set of valid and reliable surveys that ask consumers and patients to report on and evaluate their experiences with health care. The findings from these surveys have been used to inform consumer healthcare choices and for quality improvement activities at the facility level. The CAHPS Nursing Home Family survey is an addition to this family of products. Research and development for this survey was sponsored by the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality. The objective of the survey is to collect data that measure family members' perceptions of nursing home care.

The CAHPS Nursing Home *Family* Survey ("family survey") complements the CAHPS Nursing Home *Resident* Survey ("resident survey") in that it was not designed to provide the same information that the resident survey provides. The family survey does not ask respondents to report on the residents experience; rather, it asks respondents to report on their own experiences. In other words, the family survey does not ask the respondent to provide proxy responses for the resident's experience. This is an important distinction because research has demonstrated that queries of family members do not provide the same data as queries of residents on nursing home satisfaction and experience-of-care surveys.

Description of Tasks and Report Structure

This report describes the ten step process to develop the survey grouped into the following four activities:

1. **Formative research.** Formative research activities included three activities; publishing a call for measures in May 6, 2004, a literature review, and focus groups with family members of nursing home residents.

2. **Development of items and cognitive testing**. The next two activities included drafting an initial set of survey items and two round of cognitive testing with family with family members of nursing home residents.

3. **Technical Expert Panel and refinement of instrument.** The next two steps were to hold a Technical Expert Panel (TEP) and refine the survey as a result. The TEP met on November 2, 2005 to review candidate items and composites. The TEP included representatives from the nursing home industry, regulators and quality improvement organizations, consumers, providers, and long term care researchers.

4. **Pilot survey, analysis, TEP review, and finalization of instrument**. The instrument was field tested in east Texas. The Texas State Long Term Care Ombudsman recruited 15 Texas nursing homes and the Health Quality Council. The CAHPS Consortium conducted the Texas field test. In addition, the Health Quality Council of Alberta conducted an independent field test in Alberta and contributed their data to AIR for the psychometric analysis conducted by AIR. The TEP provided a review after wards to ensure that necessary substantive items were included.

In the next sections, we describe each of these four activities.

I. Formative Research

The formative research included focus groups, a call for measures, and reviewing literature.

Focus Groups

The goals of the focus groups were to:

- Understand participants' current experiences with nursing homes.
- Determine how participants' conceptualize good care.
- Determine the comparative salience and importance of the factors associated with good care.
- Understand participants' potential uses of a nursing home quality report.

In order to answer these questions, the three CAHPS grantees (Harvard University, AIR, and RAND, conducted 12 focus groups: two each in New York City and Phoenix, Arizona and four each in Palo Alto, California and Chapel Hill, North Carolina. With the exception of those in North Carolina, all focus group participants had already chosen a nursing home for a family member. Those participants in North Carolina were at the stage of considering moving a relative to a nursing home within a year. Findings from these focus groups can be found in ***Appendix A, Focus Group Findings from AIR, RAND, and Harvard.***

The call for measures and the literature review were conducted as part of the initial Nursing Home *Resident* Survey and are not included here.

II. Development of Items and Cognitive Testing

Item Development

We used the results from the call for measures, the literature review, and the focus groups, to develop a list of initial domains and questionnaire items within each domain (see ***Table 3. Draft Domains and Item Handles Prior to Cognitive Testing***).

Table 3. Draft Domains and Item Handles Prior to Cognitive Testing

Domains and Item Handles
1. Domain: Communication
• Converse with staff about family member's care
• Courtesy and respect
• Language differences
• Obtaining information about family member's medical condition
• Nurses/aides speak to you about how resident doing/feeling
• Comfort with asking questions about family member
• Getting information about payments and expenses
• Management availability
• Staff ability to manage concerns if there were concerns
• Concern about staff taking out your issues on family member
• Stop yourself from talking to staff because of concern about staff taking it out on resident
2. Domain: About Staff Behavior (psychosocial)
• Really care about resident
• Treat resident with courtesy
• Treat resident with kindness
• Staff checking on residents
• Staff rude to residents
• Staff treat residents roughly
• Family member ever complained to you about being physically abused or treated roughly by staff?
• Family member ever complained to you about being physically abused or treated roughly by other residents?
• Waiting for eating because of staff delay
• Waiting for drinking because of staff delay
• Waiting for dressing, toileting, or bathing because of staff delay
3. Domain: About the Nursing Home (Environment)
• Problems with laundry service (if applicable)

•	Cleanliness of room
•	Working order of equipment
•	Acceptability of noise level near room
•	Ability to find places to talk in private
•	Cleanliness of public areas
•	Unpleasant smells
4.	**Domain: Care of Your Family Member**
•	Family member look and smell clean
•	Staff help family member when needed
•	Effort to be gentle by aides and nurses
•	Nurses and aides tell family member what to expect
•	Protect modesty
•	Encouraging of family member to participate in care decisions
•	Encourage family member to be independent
•	Same nurses and aides provide care to family member
•	Enough involvement in care
•	Staff consult you about family member's care
•	Participation in care conference
•	Concern about food intake
•	Concern about hydration
5.	**Ratings**
•	Keeping track of family member's belongings
•	Ability to make room home like
•	Public areas homelike
•	Availability of nurses and aides
•	Care given by nurses and aides
•	Care given by nurse practitioners and doctors
•	Management
•	Overall nursing home rating
•	Recommendation

Cognitive Testing

We prepared a formatted "cognitive testing" version of the instrument and a cognitive interviewing protocol for use by the interviewers. This protocol provided a listing of scripted probes that could be employed to provide insights into each respondent's cognitive processes as he or she read and answered the pilot items. It also included a series of general questions

about the items, to allow the respondent to provide additional feedback about the items and to help assess the comprehensiveness of the instrument. A think aloud training exercise, with practice questions and a scripted response for the interviewers to use in modeling appropriate thinking aloud behaviors, was also included (see ***Appendix B: Draft Survey and First Cognitive Testing Protocol***).

The cognitive interviews examined the following issues related to the draft questionnaires:

1. **Content**: Are the questions that are included in the survey important to consumers? Are consumers able to make judgments about the questions?

2. **Comprehension and Interpretation**: Are the words, phrases, and questions easy to understand regardless of education level or knowledge of nursing home care? Are the questions interpreted as intended?

3. **Recall:** Are consumers able to recall the events asked about and to make judgments about them?

4. **Navigation:** Did the survey flow correctly? Were people skipped out of sections appropriately?

The results of the first round of cognitive testing are included in ***Appendix C: First Round Cognitive Testing Memos.*** Following the first round of cognitive tests, we revised the items and prepared a protocol for the second round of cognitive testing (see ***Appendix D: Draft Survey and Second Cognitive Testing Protocol***). This round tested the items as both self-administered items and as interviewer-administered items under the assumption that the final instrument would probably be administered by both mail and telephone. The results of the second round of testing are included in ***Appendix E: Second Round Cognitive Testing Memos.*** The instrument was revised again after the second round of cognitive testing.

Anyone with a family member in a nursing home was eligible to participate in the cognitive interviews; however persons were selected so as to assure variation in race, ethnicity, and education. The team conducted a total of 27 interviews in the first testing round in June 2005 and conducted another 27 interviews in the second round in June 2006.

Trained cognitive interviewers conducted one-on-one, in-person interviews. Using a "think-aloud" approach for the interviews, we asked the participant to read each question (or read the

question verbally to the participant), provide a verbal response, and explain the reason for the response. We then followed up with probes after each of the questions to ask about specific potential problems with each item. Immediately after each interview, the interviewer wrote a summary of the participant's comments. A member of each grantee team reviewed these summaries and extracted common themes for each item. At the end of each round, the team met to review these themes and make recommendations to the larger family survey team. The larger team made decisions together regarding each of the items.

III. Technical Expert Panel and refinement of instrument

Technical Expert Panel

Between the two rounds of testing, the team met with nursing home advocacy organizations to obtain their guidance and suggestions about the draft survey. The Technical Expert Panel included representatives from the following organizations:

- AARP
- American Health Care Association
- National Network of Career Nursing Assistants
- American Association of Homes and Services
- PIOSC: Quality Partners of Rhode Island
- Veterans Administration
- National Citizen's Coalition for Nursing Home Reform
- Scripps Gerontology Center of Miami University
- Alzheimer's Association
- American Medical Directors Association
- Gerontology Program of Towson University

Refinement of instrument

Based on the initial cognitive testing findings and suggestions, we made revisions to the instrument and tested the instrument again. After the second round of cognitive testing, the instrument underwent additional revisions (see ***Appendix E: Second Cognitive Testing Memos***). The final domains and item handles are listed in ***Table 4, Draft Domains and Items Used for Pilot Study***, below. The survey instrument that we produced following both rounds of cognitive testing is included as ***Appendix F, Pilot Study Survey*).**

Table 4. Draft Domains and Items Used for Pilot Study

Domains and Item Handles
1. Domain: Getting Care Quickly
• In the last 6 months, during any of your visits, did you try to find a nurse or aide for any reason? Yes/No (**Y/N**)
• In the last 6 months, how often were you able to find a nurse or aide when you wanted one? Never, Sometimes, Usually, or Always (**NSUA**)
• In the last 6 months, during any of your visits, did you help your family member with eating? Y/N → skip
o Was it because the nurses or aides either didn't help or made him or her wait too long? Y/N
• In the last 6 months, during any of your visits, did you help your family member with drinking? Y/N→ skip
o Was it because the nurses or aides either didn't help or made him or her wait too long? Y/N
• Help toileting includes helping someone get on and off the toilet or helping change disposable briefs or pads. In the last 6 months, during any of your visits, did you help your family member with toileting? Y/N →skip
o Was it because the nurses or aides either didn't help or made him or her wait too long? Y/N
• In the last 6 months, how often did you feel there were enough nurses and aides in this nursing home? NSUA
2. Domain: Quality of Care by Nurses and Aides
• In the last 6 months, how often did you see the nurses and aides treat your family member with courtesy and respect? NSUA
• In the last 6 months, how often did you see the nurses and aides treat your family member with kindness? NSUA
• In the last 6 months, how often did you feel that the nurses and aides really cared about your family member? NSUA
• In the last 6 months, did you ever see any nurses or aides be rude to your family member or any other resident? Y/N
• In the last 6 months, did you ever see any nurses or aides treat your family member or any other resident roughly? Y/N
• In the last 6 months, how often did your family member look and smell clean? NSUA
• In the last 6 months, did you ever see the nurses and aides fail to protect any resident's privacy while the resident was dressing, showering, bathing, or in a public area? Y/N
• Sometimes residents make it hard for nurses and aides to provide care by doing things like yelling, pushing, or hitting. In the last 6 months, did you see any resident, including your family member, behave in a way that made it hard for nurses or aides to provide care? Y/N→ skip
o In the last 6 months, how often did the nurses and aides handle the situation in a way that you felt was appropriate? NSUA
3. Domain: Communication of Nurses and Aides
• In the last 6 months, how often did the nurses and aides treat you with courtesy and respect? NSUA
• In the last six months, did you want to get information about your family member

from a nurse or an aide? Y/N → skip
    - In the last 6 months, how often did you get this information as soon as you wanted? NSUA
- In the last 6 months, how often did the nurses and aides explain things in a way that was easy for you to understand? NSUA
- In the last 6 months, did the nurses and aides ever try to discourage you from asking questions about your family member? Y/N

4. Domain: Communication--Other Staff and Administrators

- In the last 6 months, were you ever unhappy with the care your family member received at the nursing home? Y/N → skip
    - In the last 6 months, did you talk to any nursing home staff about this? Y/N → skip
    - In the last 6 months, how often were you satisfied with the way the nursing home staff handled these problems? NSUA
- In the last 6 months, did you ever stop yourself from talking to any nursing home staff about your concerns because you thought they would take it out on your family member? Y/N
- In the last 6 months, have you been involved in decisions about your family member's care? Y/N→ skip
    - In the last 6 months, how often were you involved as much as you wanted to be in the decisions about your family member's care? NSUA
- In the last 6 months, did you need to contact the nursing home administrator about any problems? Y/N→ skip
    - In the last 6 months, how often were you satisfied with the way he or she handled these problems? NSUA
- A care conference is a formal meeting about care planning and health progress between a care team and a resident and his or her family. In the last 6 months, have you been part of a care conference? Y/N
- In the last 6 months, did you ask the nursing home for information about payments or expenses? Y/N → skip
    - In the last 6 months, how often did you get all the information you wanted from the nursing home about payments or expenses? NSUA
- In the last 6 months, has any nursing home staff told you that they could not give you information about your family member because of privacy laws? Y/N

5. Domain: Nursing Home Environment

- In the last 6 months, how often did your family member's room look and smell clean? NSUA
- In the last 6 months, how often was the noise level around your family member's room acceptable to you? NSUA
- In the last 6 months, how often were you able to find places to talk to your family member in private? NSUA
- In the last 6 months, how often did the public areas of the nursing home look and smell clean? NSUA
- In the last 6 months, did you ever see a resident be rude to or treat your family member or any other resident roughly? Y/N
- Personal medical belongings are things like hearing aides, glasses, and dentures. In the last 6 months, how often were your family member's personal medical belongings damaged or lost? Never/once/two or more times
- In the last 6 months, did your family member use the nursing home's laundry service

for his or her clothes? Y/N→ skip
o In the last 6 months, when your family member used the laundry service, how often were clothes damaged or lost? Never/once or twice/3+ times
6. Ratings
• Using any number from 0 to 10 where 0 is the worst care possible and 10 is the best care possible, what number would you use to rate the care at this nursing home? 0-10
• If someone needed nursing home care, would you recommend this nursing home to them? Definitely no/probably no/ probably yes/ definitely yes

IV. Pilot survey, analysis, TEP review, and finalization of instrument

Pilot Survey

The pilot survey was conducted in Texas. Health Quality Council of Alberta had its own field test and supplied data to AIR for psychometric analyses. The Canadian instrument was slightly modified as a result of differences in the organization of nursing home care in Canada and the U.S. The following description of the methods focuses on the Texas pilot study.

Sampling Frame

Nursing Home Recruitment

AIR worked with its subcontractor, Texas A&M University, to identify a diverse range of nursing homes for the pilot survey. We were assisted by the Texas Ombudsman Office within the Texas Department of Aging and Disability Services. Ombudsmen recruited 18 nursing homes from the Dallas and East Texas regions. We recruited Medicare and Medicaid certified nursing homes that represented a range of quality scores in the Texas Quality Reporting System (**QRS)**. Nursing facilities that accept Medicaid or Medicare are compared in the QRS on the basis of four dimensions of quality: technical quality, quality of life, regulatory compliance, and customer satisfaction. The mean score on the 1-100 scale is 56. Of the 20 nursing homes that initially agreed to participate, 15 nursing homes actually participated in the survey. Of the five nursing homes which opted out, reasons for declining included: 1) concern for respondent burden (they limit number of family surveys to one every 6 months); 2) the nursing home was closing down within a month; 3) low number of eligible patients; 4) change in key nursing home staff; and 5) unable to provide contact information for family members in time for survey.

Of the 15 participating nursing homes, 5 had QRS scores below 56, the average, one facility had a score of 56, and the other 9 exceeded the average. The average rating of participants was 62.6, slightly above the overall statewide mean of 56. The average rating for the lower-scoring nursing homes is 30.2 and the average rating for the higher-scoring nursing homes was 81.4 (not including the nursing home scored at 56.

Sampling Frame Eligibility

Eligible sample members were identified first by identifying residents who met the eligibility criteria and then identifying the responsible person for those residents. An eligible sample member was the person listed by the nursing home as the responsible party for a resident who had resided at the facility for at least 30 days. In addition to family members and friends, guardians, medical powers of attorney, and attorneys were considered to be eligible responsible parties. Once we received the list of eligible responsible parties from the nursing home, we excluded or re-designated the eligible responsible party when, in our judgment, the named party would have been uninformed or biased about the nursing home's performance.

1. In 31 cases, representing 2.1% of the responsible parties received from the nursing homes, the responsible party was a nursing home resident (including self) or a nursing home staff member or a trust fund or another institution. All of these were excluded.

2. In 28 cases (1.9%), the responsible party listed was the resident or the nursing home, but there was another, eligible contact listed as well. In these cases, the other contact was used instead of excluding the case.

3. In 44 cases (3.0%), residents had more than one eligible responsible party listed. This does not include the other contacts most residents had listed; only cases in which multiple persons were listed as responsible parties for the same resident. In these cases, one person was randomly selected to participate in the survey.

4. Another 17 eligible sample members (1.2%) were listed as the responsible party for more than one resident. For these cases, we randomly selected a resident for them to respond about.

5. In 12 cases (0.8%), an eligible responsible party was listed without contact information, but another person listed as a contact (not a responsible party) had adequate contact information. In these cases, we used the alternative with adequate contact information.

In addition to the above exclusions based on administrative data, the questionnaire included several screening questions that excluded other cases. It screened out eligible sample members who had visited the focal resident less than once in the last six months, and whose

focal resident had been discharged from the focal nursing home at the time of the survey. Participants with family members who were discharged were excluded, because it became apparent during cognitive testing, that these were two different audiences and a separate instrument was needed to examine experiences of people who had a nursing home resident recently discharged or transferred from a nursing home.

Sampling Methodology

Each participating nursing home forwarded a data file containing the following items for each eligible respondent: name of responsible party, address, phone number, resident/patient name, date-of-birth, gender, relationship of responsible party to resident, whether responsible party was the power of attorney, admission date, and whether the resident was in a dementia unit.

Our goal was to sample 150 potential respondents per nursing home, but the average number of eligible frame members was approximately 90 per nursing home. When the frame contained fewer than 150 eligible sample members, we sampled everyone. In one case, the number exceeded 150 by a small amount and we also sampled everyone. The initial sample size was 1,471; after removing those without an address, the sample fell to 1,444. The survey began in late October 2006 and ended early January 2007. All eligible respondents were mailed the family survey questionnaire, with a cover letter and a return postage-paid envelope. A reminder postcard was sent approximately 2 weeks after the first questionnaire was mailed. A second questionnaire and cover letter was sent approximately 2 weeks after the postcard was sent. Approximately 2 weeks after that, computer-assisted telephone interviews (CATI) began for nonrespondents. A maximum of five telephone attempts were made on different days and times of the day to try to maximize response rates. The final response rate was 66 percent (N = 885). Response rates differed by wave and mode:

- Wave 1 (initial survey): 42%
- Wave 2 (second survey): 14%
- Phone follow-up: 10%

Analysis

The AIR team, with review by RAND and Yale team members, conducted multiple activities to analyze the data for validity, reliability, and case mix. The analysis included the data from the 14 nursing homes in the Alberta pilot survey that had taken place two months earlier. We generated descriptive statistics at the item level and identified missing data, out of range values, and skip pattern inconsistencies (***Appendix G, Frequencies from pilot study)***. We also generated descriptive statistics presenting the Texas and Alberta data. The instrument team met and evaluated each item to determine whether they were similar enough to be used together analytically. In most cases, and in all cases of critical content, we found a high level of convergence at the item level.

Item Functioning

Survey questionnaires that are poorly designed – that is, those with complicated skip patterns, hard-to-understand instructions, a readability level that is beyond many respondents, or that are otherwise too complex – may yield unreliable results. We can assess how well designed a survey is by observing how well the items on the survey 'function.' The functioning is measured by analyzing patterns of inconsistent or missing responses to survey items, such as a respondents' failure to follow skip instructions or the relative amount of missing data.

The family survey contained several skip patterns; each skip consists of a gate, or screener, and one or more items controlled by that gate item. To assess how well the survey functioned, we examined every gate-item pair (i.e., skip) in the survey, including nested skips. We also examined any items that were not part of a skip pattern to check for missing data – in all, we assessed 74 items (Q02 through Q66)[1] and assigned one of five dispositions to every item, for every respondent:

1. Correct Skip (CS) – the respondent (correctly) skipped an item that they were supposed to skip.

[1] Note: the total number of items is greater than the number of questions because some items, such as race (Q63) and proxy (Q66), are 'code-all-that-apply' items, so a single item actually produces multiple variables, each of which indicates whether or not the respondent marked one of the response options.

2. Failed Skip (FS) – the respondent answered an item that they should have skipped.
3. Indeterminate Eligibility (IE) – the respondent left the gate preceding the item blank, so eligibility to answer the item cannot be determined.
4. Truly Missing (TM) – the respondent was eligible to answer the item, but left it blank.
5. Legitimate Response (LR) – an eligible respondent gave a legitimate answer to the item.

For each item, we calculated the percentage of respondents who fell into each of these five dispositions. Table 5 provides and example of the distribution of these dispositions for question 11 on the survey:

Table 5. Item Functioning Dispositions for Q11

Disposition	Frequency	Percent
Correct Skip	141	16.53
Failed Skip	32	3.75
Indeterm Elig	13	1.52
Truly Missing	8	0.94
Legit Resp	659	77.26
Total	**853**	**100.00**

We focused in particular on the percentage of respondents who had failed skips and the percentage of respondents with missing responses for each item to identify items that respondents may have had trouble completing. High rates of missing data or skip failures for an item may indicate that the question was confusing, too personal, or offensive. Our analysis showed that the rate of truly missing was less than 5% for all survey items that were intended to be used for reporting (Q04 – Q51). The rate of failed skips was less than 10% for all survey items that were part of a skip pattern.

We calculated some summary measures for the item dispositions by creating count variables that sum each occurrence of each of the five dispositions across all 853 respondents. For example, the number of items an ineligible respondent failed to skip is summed for each respondent, then that total is summed across all 853 respondents, and the mean calculated.

The same is done for the other dispositions. Out of 74 survey items, on average, respondents made fewer than one skip error, had slightly more than two truly missing responses, and had slightly more than two items where we could not determine their eligibility to respond.

We also calculated an item-level response rate (IRR), which, for each respondent, is equal to the number of items with a legitimate response divided by the number of items that the respondent was eligible to answer.

The rate of item non-response (RINR) is the complement to the IRR, or:

$$RINR = 1 - IRR$$

The mean item-level response rate for all 74 items across all respondents was 93.6%.

Table 6 gives the mean item-level response rate by nursing home and the percent of respondents at each facility who had non-missing data for at least 90% of the items they were eligible to answer (i.e., their rate of item non-response was 10% or less).

Overall, item nonresponse for this survey instrument was comparable to what was found with the pilot CAHPS Hospital survey, which was in the range of 2 to 4 percent (Elliott, Edwards, Angeles, Hambarsoomians & Hays, 2005).

Table 6. Mean Item-Level Response Rate by Facility

Nursing Home Number	Mean Facility-level IRR	% Respondents with IRR of 90% or Greater
1	91.3%	75.6%
2	94.8%	90.0%
3	93.5%	90.0%
4	89.9%	72.1%
5	96.1%	94.3%
6	92.1%	85.2%
7	96.4%	93.8%
8	92.0%	77.8%
9	93.9%	82.5%
10	94.2%	93.1%
11	93.6%	84.4%
12	94.9%	91.2%
13	92.2%	85.7%
14	93.2%	80.0%
15	91.0%	76.9%
Total	**93.6%**	**86.3%**

The combined problem of missing data and skip errors was a noticeable problem in one section of the survey that includes questions that ask about legal matters related to the residents care (questions 56, 57 and 58[2]), none of which are used in the composites. The rate of item nonresponse was 24% for Q56 and 23% for Q58, and the rate of failed skips for Q57 was over 7%.

[2] Q56: A court-appointed legal guardian has the authority to make decisions for another person because the other person cannot make decisions. Are you your family member's legal guardian? [if Yes, Go to Question 58]
Q57: Is someone else your family member's legal guardian? [Yes, No, Don't Know]
Q58: In the last 6 months, has any nursing home staff told you that they could not give you information about your family member because of privacy laws? [Yes, No]

Data Cleaning

We followed standard CAHPS data cleaning rules:

1. For gate-item pairs where the gate was a yes-no question and the item was a 'how often' question: If the gate was left blank, and the follow-up item was answered 'never', we coded the follow-up as missing. If the follow-up was answered SUA (sometimes, usually, or always), we kept the response to the follow-up, and back-coded the gate question to 'Yes'.

2. If a gate question was missing (blank/not ascertained), and subsequent survey items controlled by that gate question contained valid responses, the responses for those items were retained.

3. Failed skips: If the response to a gate question was valid, but the respondent violated the skip instruction by answering survey items that should have been skipped, the response to the gate question was retained and the responses for survey items within the gate were set to missing.

4. Correct skips were coded as missing.

Psychometric Analysis

Prior to data collection and in consultation with stakeholders, we organized 31 of the substantive survey items into a set of five domains. The first step in the psychometric analysis was to test whether there was empirical evidence to support the hypothesized item-domain relationships. We combined the Alberta and Texas data, and used that combined data set to calculate a Cronbach's alpha for each of the five domains. The alphas for 3 out of the 5 domains were greater than the standard 0.70, and the alphas for two of the domains were lower. While these results provide some support for the hypothesized relationships, the item-total correlations for the domains indicated that each domain's alpha would improve if certain items were dropped, which indicates that it was likely that there was a better item-domain

structure to be specified. Separate analyses of the Alberta and Texas data replicated these findings.

Our next step was to make use of some exploratory methods to identify composites. To make use of all available data, we obtained maximum likelihood estimates of the covariance matrix under the Missing at Random (MAR) model using a multiple imputation procedure (MI, Rubin 1976, 1987).[3] Peer-reviewed publications involving another CAHPS instruments illustrate the use of the MAR model and SAS PROC MI for this purpose (Hurtado, Angeles, Blahut & Hays, 2005; Keller et al., 2005; O'Malley et al., 2005).

We conducted an exploratory factor analysis (EFA) on the correlation matrix produced by PROC MI. The EFA used the principle factor method with squared multiple correlations as initial communality estimates, and oblique rotation (promax) with Kaiser normalization. The number of factors was determined by the eigenvalues and the interpretability of the rotated factor pattern matrix. The EFA results did not demonstrate a definitive underlying factor structure for these 31 items. The analysis proceeded at this point through an iterative process that included additional factor-analysis and multi-trait multi-item analysis, along with conversations among various members of the analysis team and CMS.

The team agreed on a final set of 21 items organized into four domains. The first three composites are more specific in focus and the fourth, more general. The first composite refers to whether respondents perceive that nurses and aides provide help with basic needs of residents who require that type of help. Only those respondents who helped their family member with eating, drinking, or toileting were eligible to answer the questions in the first composite (approximately 30% of the respondents). As might be expected, the residents that these respondents visited tended to be more impaired than the residents visited by respondents who did not help with eating, drinking and toileting (data not reported but available upon request). For convenience, we refer to this composite as "Meeting Basic Needs."

[3] The MAR model and SAS PROC MI in particular has been recommended as a method of producing a complete data file preliminary to analyses on the underlying structure of questionnaires (Tomarken & Waller, 2005). By producing multiple estimates for each missing data value, MI does not restrict the total variance in the data matrix as does a single imputation procedure.

The second composite refers to the interpersonal manner in which the nurses and aides interact with residents, "Nurse/Aide Kindness", for short. The third composite refers to how much the nursing home shares information related to the resident's care with the respondent, or "Nursing Home Information." Finally, the fourth composite contains a variety of items that speak to the general quality of the care delivered by the nursing home. For convenience we call it "Nursing Home Care Quality," although the content refers to many specific aspects of care. See Table 3 on page 19 for the list of composites and related items.

A series of analyses were conducted to determine the measurement properties of the items using the US data. Results are reported with the US data because the survey is most likely to be used in the US; however, similar results were obtained with the Canadian data. Table 7 summarizes the disposition of the 31 items tested including the composite structure of the 21 items that were retained along with their measurement characteristics. The last column of the table explains why two of the items were kept despite having marginal measurement characteristic and why each of the 10 items was dropped.

As shown in the table, all four composites demonstrate sufficient criterion validity, as evidenced by their relatively high correlations (> 0.30) with the three global measures. Although the observed facility-level reliability of the first composite – Meeting Basic Needs – is not as high as we would like, it will be able to discriminate across nursing homes, given a sufficient number of respondents per facility. As mentioned above, this composite has a high percentage of missing data because it is made up of three items that were appropriately skipped by a large number of respondents.

Table 7. Final Composites and Dispositions

		Correlation with Globals (> .30)			NHr[c]	
Q#	**Composite or Item Text**	**Ever Unhappy**	**Recom-mend**	**0-to-10 NH Rating**	**> .70**	**Justification for Retaining/Dropping Item**
	Meeting Basic Needs: Help with eating, drinking, toileting[a]	**0.44**	**0.46**	**0.54**	**0.48**	
19	Wait too long for help with eating	0.40	0.44	0.50	0.46	
21	Wait too long for help with drinking	0.33	0.36	0.41	0.32	
23	Wait too long for help with toileting	0.57	0.51	0.56	0.55	
	Nurses/Aides' Kindness/ Respect Towards Resident	**0.52**	**0.59**	**0.70**	**0.83**	
12	Nurses/Aides treat resident with respect	0.36	0.47	0.55	0.77	
13	Nurses/Aides treat resident with kindness	0.43	0.48	0.59	0.81	
14	Nurses/Aides really cared about. resident	0.49	0.55	0.70	0.82	
15	Nurses/Aides rude to resident	0.39	0.38	0.41	0.64	
26	Nurses/Aides appropriate with violent resident	0.41	0.50	0.62	0.16	
	Nursing Home Provides Info/ Encourages Respondent Involvement	**0.50**	**0.64**	**0.66**	**0.85**	
29	Nurses/Aides give respondent information about resident	0.48	0.54	0.60	0.85	
30	Nurses/Aides explain things to respondent	0.45	0.53	0.56	0.78	
31	Nurses/Aides discourage respondents questions	0.19	0.26	0.24	0.48	Marginal measurement characteristics. But Focus Groups and Ombudsmen liked; Scaled well with its composite.
43	Respondent stops self from complaining	c	0.43	0.38	0.50	
45	Respondent involved in decisions about care	0.33	0.44	0.48	0.75	
53	Respondent given Info. about payments/expenses	0.32	0.57	0.51	0.63	
	Nursing Home Staffing, Care of Belongings, & Cleanliness	**0.55**	**0.67**	**0.77**	**0.89**	
11	Can find a nurse or aide	0.40	0.54	0.57	0.76	
51	Enough nurses/ aides?	0.46	0.57	0.64	0.88	
32	Room looks/smells clean	0.41	0.49	0.57	0.82	
24	Resident looks/ smells clean	0.42	0.44	0.55	0.85	
35	Public areas look/smell clean	0.37	0.52	0.58	0.85	

Table 7. Final Composites and Dispositions

		Correlation with Globals (> .30)			NHr[c]	
Q#	**Composite or Item Text**	**Ever Unhappy**	**Recom-mend**	**0-to-10 NH Rating**	**> .70**	**Justification for Retaining/Dropping Item**
37	FM's medical belongings lost	0.21	0.22	0.24	0.21	Marginal measurement characteristics. But Focus Groups and Ombudsmen liked; Scaled well with its composite.
39	FM's clothes lost	0.29	0.35	0.40	0.82	
Items Dropped from Composites						
16	Nurses/Aides treat resident rough	0.24	0.28	0.32	0.14	Marginal measurement characteristics. Group agreed to drop.
17	Another resident rude to respondent's family member	0.20	0.21	0.26	0.08	Marginal measurement characteristics. Group agreed to drop.
27	Nurses/Aides treat respondent with respect	0.41	0.51	0.54	0.74	Not uniquely and strongly related to its composite. Other items tap concept. Group agreed to drop.
33	Noise level around room	0.27	0.38	0.41	0.79	Not uniquely and strongly related to its composite. Group agreed to drop.
34	Places to talk to resident in private	0.26	0.31	0.35	0.76	Not uniquely and strongly related to its composite. Other items tap concept. Group agreed to drop.
36	Nurses/Aides didn't protect resident's modesty	0.20	0.23	0.23	0.56	Marginal measurement characteristics. Group agreed to drop.
42	Staff handled respondent's concerns satisfactorily	c	0.67	0.68	0.68	Includes term "satisfaction" which is not consistent with CAHPS methods; also low NH-reliability
46	Nursing Home has care conferences	0.00	-0.20	-0.15	0.82	Care conferences are required in Canada ONLY.
48	Management handled complaints satisfactorily	0.41	0.61	0.62	0.59	Includes term, "satisfaction" which is not consistent with CAHPS methods; also low NH-reliability
58	Staff did not give respondent info. due to privacy laws	-0.02	0.00	0.01	0.0	Marginal measurement properties. Group agreed to drop.

[a] Composite labels are in bold font. Composite scores are calculated as the mean the scores of the items that make up the composite. Composite score was only calculated for respondents who had non-missing data for half or more of the items that make up the composite.

[b] Nursing Home Reliability, or Inter-unit reliability (IUR). This statistic represents a transformation of the F-statistic for testing differences among nursing homes on an item or composite (IUR = (F - 1)/F). IUR can be interpreted as the fraction of the variation among nursing home scores that is due to real differences rather than chance. The higher the IUR, the greater the ability of the item or composite to discriminate across nursing homes. An IUR > 0.70 is considered to indicate a high level of discriminant ability for an item or composite.

[c] Q40 is the screener item that controls whether or not the respondent skips Q42 and Q43. Only those who responded 'yes' to Q40 were eligible to respond to either Q42 or Q43

Table 8 presents the composite-level psychometrics for the four final composites in the US data. The internal consistency reliability (Cronbach's alpha) was relatively high for all four composites, which indicates that the scores would provide reliable data. With the exception of the Nursing Home Care Quality composite, the item-total correlations and scaling success of the composites were all high. With regard to the Nursing Home Care Quality composite, these results taken together indicate that the composite has reliable scores, but overlaps in meaning with some content in the other composites. However, this finding is to be expected given that the composite is a general indicator of nursing home quality and actually indicative of the composite's validity as a more general measure. The Nursing Home Information and Nursing Home Care Quality composites both contain one item each that was weakly related to the overall total score, corrected for overlap (data not reported but available upon request). Both of these items were retained in the survey in response to the concerns of potential respondents and ombudsman and are the same items flagged in Table 7 (i.e. Q31 and Q37).

Table 8. Composite-level Psychometrics

Composite	Mean (SD)	Alpha	Median Correlation with Own Composite[a]	% Missing Data[b]	Scaling Success (%)	% at Floor	% at Ceiling
Meeting Basic Needs: Help with eating, drinking, toileting	73.5 (39.3)	0.90	.80	1.8%	100%	19%	64%
Nurses and Aides' Kindness and Respect Towards Family Members	84.8 (19.9)	0.88	.77	3.1%	100%	0%	9%
How Well the NH Provides Info and Encourages Family Involvement	87.4 (17.0)	0.78	.58	9.1%	100%	0%	15%
NH Staffing, Care of Belongings, and Cleanliness	80.5 (17.1)	0.79	.58	4.7%	86%	0%	9%

[a] Correlation between item and the composite to which it belongs, corrected for overlap. Within each composite, these correlations are indicators of convergent validity, and should, as a rule of thumb, be greater than 0.40.
[b] % Missing denotes the percentage of eligible respondents for whom the composite could not be scored without imputing missing data for the items within the composite.

The Meeting Basic Needs composite was limited in the amount of information it could provide because of limited degrees of freedom. That is, due to the dichotomous response format, there were only 5 possible scores. Thus, more than 60% of eligible respondents provided answers indicative of the highest quality care, while another 20% provided answers indicative of the

lowest level of quality of care on that composite. Variability for this composite could be improved by including more items. However, we decided not to add to the composite at this time in order to minimize respondent burden and because some stakeholders argued against the inclusion of the composite because they perceived the content as too critical of nursing homes. The other three composites demonstrated a great deal of variability in scores – very few of the scores observed were either the highest possible (at the ceiling) or lowest possible (at the floor).

Overall all four composites demonstrated good psychometric properties, as shown in the Table 9. The Nursing Home Care Quality composite lacks some discriminant validity due to its relationship to other composites. The basic needs composite has relatively low facility-level reliability, but that can be partially remedied with larger sample sizes.

Table 9. Summary of Psychometric Criteria

	Met Criterion?			
Statistical Criterion	**Basic Needs**	**Staff Kindness and Respect**	**NH Info and Encouragement**	**NH Staffing, Care, Cleanliness**
NH-Level Reliability (> 0.70)	No	Yes	Yes	Yes
Internal Consistency Reliability (> 0.70)	Yes	Yes	Yes	Yes
Criterion Validity (> 0.30)	Yes	Yes	Yes	Yes
Convergent Validity (> 0.40)	Yes	Yes	Yes	Yes
Scaling Success (~100%)	Yes	Yes	Yes	No

Case-mix Analysis

Background

In order to compare facilities included in this study, it is important to control for the influence of respondent characteristics on the outcome variables. Past research on health plans has shown that some types of respondents, such as older respondents, or those who are in better health, tend to give more positive responses to CAHPS items than other types of respondents (O'Malley, Zaslavsky, Elliott, Zaborski, & Cleary, 2005). Conversely, those respondents with more education or poorer health tend to give less positive responses to CAHPS items. These are characteristics of the respondents that are related to the CAHPS scores, but do not constitute characteristics of the facility, system, or plan being evaluated. Nor are they believed to reflect true differences in the quality of care delivered.

Generally speaking, when comparing facilities, the differences should derive entirely from differences in the quality of care they provide. The family survey items are designed to achieve this goal. However, if facility differences derive in part from differences in the respondent or resident populations in those facilities, rather than entirely from differences in the quality of care they deliver, it is important to remove the portion of the scores that come from respondent/resident characteristics so that comparisons accurately reflect differences in quality. These differences in the populations are called *case-mix* and removing their contribution to the scores is called *case-mix adjustment*.

The three basic goals of case-mix adjustment in the analysis of patient assessments of care are to:

1. Help remove the effects of individual patient characteristics that can affect patient or family member experiences,
2. Remove effects that might be considered spurious (i.e., that reflect something other than quality of care), and
3. Remove incentives for facilities to avoid enrollment of "hard to treat" patients (Zaslavsky, 1998).

Zaslavsky (1998, p.58) outlines three conditions to be met in the selection of variables for case-mix adjustment.

1. Within the facilities, the case-mix variables must be related to the outcome measures (ratings). That is, the variables must have sufficient *predictive power* in relation to the outcomes (e.g., older respondents give higher ratings of the nursing home care).

2. There must be variation between facilities on the case-mix variables. That is, the variables must be unevenly distributed across facilities (e.g., one facility might have a significantly higher percentage of elderly family respondents than another). This condition is the *heterogeneity* of the predictor.

3. The case-mix variables must be appropriate for adjustment because they are not themselves determined by the facilities' actions. That is, they must be characteristics that are brought to the facility by the respondent or patient (e.g., age or education), not characteristics that might be consequences of satisfaction with or assessment of the facility. For example, the length of relationship with the facility can reasonably be considered a *consequence*, rather than a *cause*, of a respondent's satisfaction with the nursing home care experiences of their family member.

This study is further complicated by the fact that the respondents are not the ones who are directly receiving the care, and are being asked to report their observations and experiences regarding the care received by someone else – the family member who is the actual resident in a nursing home facility. The initial choice of potential case-mix adjusters reflects an effort to take into consideration both the characteristics of the respondent (e.g., the respondent's age, gender, education, and number of times the respondent visits the resident) and the resident. One measure typically used as an adjuster for CAHPS analysis is the respondent's self-rating of their overall health. We have no such measure of the respondent's health on this survey; but we do have analogous measures of the resident's health (e.g., question items about the resident's capacity to make their own decisions or the resident's memory problems).

The case-mix analysis proceeds through the following steps:

1. Selection of potential case-mix adjusters
2. Estimation of predictive power of the selected adjusters
3. Estimation of heterogeneity
4. Calculation of explanatory power and impact of each adjuster

Selection of potential adjusters

We chose an initial set of adjusters to evaluate based on both historical use – some variables, such as age and education, have been subjected to extensive case-mix analysis in other CAHPS studies, and there is ample evidence that they are important case-mix adjusters – and the conceptual appropriateness of the variable as an adjuster. The next step was to select a subset of these potential case-mix adjusters for further analysis. The strength of the relationship of each potential adjuster to overall ratings of nursing home care was evaluated using step-wise regression in which each potential adjuster was regressed onto three global-type outcomes.[4] Our initial pool of potential adjusters consisted of the following:

- Respondent gender
- Respondent age
- Respondent education
- Resident has serious memory problems (q07)
- Resident ever had to share a room (q06)
- Expect resident to live in this or other Nursing Home permanently (q05)
- How often resident capable of making decisions about own daily life (q08)
- Number times in last 6 months respondent visited resident (q09) [5]

[4] The stepwise method is a modification of the forward-selection technique and differs in that variables already in the model do not necessarily stay there. As in the forward-selection method, variables are added one-by-one to the model, and the F-statistic for a variable to be added must be significant at the inclusion p-level. After a variable is added, however, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an F-statistic significant at the exclusion (retention) p-level. Only after this check is made and the necessary deletions accomplished can another variable be added to the model. The stepwise process ends when none of the variables outside the model has an F statistic significant at the inclusion p-level and every variable in the model is significant at the retention p-level, or when the variable to be added to the model is the one just deleted from it. Following O'Malley et al. we set the inclusion p-value criterion at 0.10, and the exclusion (retention) criterion at 0.05 (i.e., to stay in the model, a given variable had to be significant at $p. < .05$).

[5] Respondent's age, education, q08, and q09 were treated as continuous variables. Respondent gender, q06, and q07 were dummy-coded and entered into the model as indicator variables with a value of 1 or 0. For gender, the reference category is female; for q06 and q07, which are yes/no questions, the reference category is 'no.' Since a substantial number of respondents answered "don't know" to q05, we retained that response category and dummy coded q05 – the reference category is 'no.'

We did not use relationship items relating to financing or responsibility (Q54-57) because other items reflected the interaction better, such as visiting the resident, because of the need to decrease the size of the survey and because two of the items, Q56 and Q57 which ask about legal guardianship presented problems with missing data and skip errors.

We modeled the relationship of our potential adjusters to three global outcome measures. Thus, three stepwise regression models were calculated in which the potential adjusters listed above were regressed on each of the three following outcome variables:

- Ever Unhappy with Care your Family Member Received[6] (q40)?
- Global Rating of Care at Nursing Home (q49)
- Would You Recommend this Nursing Home (q50)?

The results of the variable-selection analysis are reported in Table 10; bolded variables met the inclusion criteria for at least one of the three outcomes.

Table 10. Case-mix Adjuster Selection

	Ever Unhappy with Care[6] (q40)	Rate Care at NH (q49)	Would Recommend NH (q50)
Potential Adjusters	*Parameter Estimate*		
Respondent Gender (Male)	0.082*	0.392*	0.124[b]
Respondent Age	0.045**	0.237***	0.077**
Respondent Education	a	a	0.082**
Resident Memory Problems (q07)	a	a	a
Resident Share a Room (q06)	a	a	a
Resident In NH Permanently Yes (q05)	a	1.167**	0.415**
Resident In NH Permanently Don't Know (q05)	a	0.821[b]	a
Resident Capable of Making Decisions (q08)	0.06***	0.131*	a
Number Times Respondent Visited Resident (q09)	-0.07***	a	a

[a] Failed to meet inclusion p-value criterion.
[b] Met inclusion criterion, but failed to meet model retention criterion.
* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

[6] It should be noted that q40 is coded so that a score with a quantitative value of one corresponds to a response equivalent to "I was **unhappy** with the care my family member received", and a score with a quantitative value of two corresponds to a response equivalent to "I was **never unhappy** with the care my family member received," so a higher score represents a more positive experience. The question label in the table has been changed to reflect that the more positive response has a higher quantitative value.

Predictive Power

The next step to identifying case-mix adjusters was to evaluate each of the remaining six variables for their relative unique strength in predicting the three outcomes. Predictive power was measured as the incremental amount of variance in each outcome explained by a potential case-mix adjuster (the predictor) in a step-wise linear regression analysis, controlling for facility and the other variables being assessed as potential case-mix adjusters. Following O'Malley, Zaslavsky, Elliott, Zaborski, & Cleary (2005), we report the power of each adjuster to predict this quality rating as the partial r^2 for that adjuster * 1,000. These values are presented in the Table 10.

Case-mix Adjuster Heterogeneity

Each potential case-mix adjuster was also evaluated for how much it differed between facilities compared to how much it differed within each facility. Heterogeneity of each potential case-mix variable (predictor) across plans was defined as the ratio of its between-facility variance to its residual within-facility variance. For case-mix adjuster heterogeneity, we ran a separate variance component model with each potential case-mix adjuster as the outcome, with the other potential case-mix adjusters specified as fixed effects, and facility specified as a random effect (i.e., we estimated the contribution of facility to the variance of the potential case-mix adjusters). These results are presented in Table 11 as well.

Table 11. Assessing the Impact of Selected Case-mix Adjusters

		Ever Unhappy with Care (q40)			Rate Care at NH (q49)			Would Recommend NH (q50)		
Outcome Heterogeneity Factor		***0.0843***			***0.2911***			***0.2763***		
Potential Adjusters	*Case-Mix Adjuster Heterogeneity Factor*	*Predictive Power*	*EP**	*Impact Factor**	*Predictive Power*	*EP**	*Impact Factor**	*Predictive Power*	*EP**	*Impact Factor**
Respondent Gender (Male)	-0.0018	6.80	-0.01	-0.15	11.00	-0.02	-0.07	8.50	-0.02	-0.06
Respondent Age	0.0107	11.30	**0.12**	**1.43**	21.60	0.23	0.79	8.90	0.09	0.34
Respondent Education	0.1587	4.30	**0.68**	**8.09**	2.90	**0.46**	**1.58**	0.40	0.06	0.23
Resident In NH Permanently (q05)	0.0212	4.60	**0.10**	**1.16**	16.90	**0.36**	**1.23**	13.00	**0.28**	**1.00**
Resident Capable of Making Decisions (q08)	0.0573	11.20	**0.64**	**7.60**	2.40	**0.14**	**0.47**	0.90	0.05	0.19
Number Times Respondent Visited Resident (q09)	0.0053	17.80	**0.09**	**1.11**	3.80	0.02	0.07	0.50	0.00	0.01

* Predictive Power = R-square*1,000; EP = heterogeneity * predictive power; Impact Factor = EP/Outcome heterogeneity
Bolded adjusters are those with an EP > 0.1 and an impact factor > 1.0

Impact

The first step in assessing the impact of each potential adjuster is to calculate the explanatory power (EP) of each variable being assessed – the product of its predictive power and its heterogeneity factor. Adjusters with an EP of at least 0.10 are considered to have the potential to have a noticeable impact on CAHPS scores (Zaslavsky, 1998). We then calculated the impact factor for each potential adjuster.[7]

An impact value of 1.0 for the case-mix variable indicates that it has the potential to result in a change in the outcome that is at least equal to the baseline variance in that score across plans. We used a threshold value of 1.0 for the impact factor to screen for potential case-mix variables, following O'Malley, Zaslavsky, Elliott, Zaborski, & Cleary (2005).

Recommendations for Case-mix Adjusters

We found statistical evidence supporting the possible use of five variables for case-mix adjusting (see bolded items in Table 10). Respondent age and education have historically been used as adjusters in the analysis of CAHPS data; in fact, they are recommended as adjusters in the documentation available on the CAHPS user web site. The results presented here provide empirical evidence for using them for the family survey analysis as well.

We also found support for the use of two of the health-rating analogues as case-mix adjusters. These were: the respondent's belief about resident's capability of making decisions and whether or not the respondent believes the resident will permanently live in a nursing home. Both exhibit a potential impact on two of the three outcomes, and the latter appears to have a reasonable impact on the recommendation outcome.

The final adjuster potentially measures the amount of 'data' available to the respondent in reporting their observations and experiences—one who visits more regularly would

[7] To calculate the impact factor, we examined the variance in the three outcome variables across the facilities (outcome heterogeneity) by estimating a variance component model for each of the three outcomes, with facility specified as a single random effect (i.e., we estimated the contribution of facility to the variance in the outcomes). This value is presented in the second row of Table 10.
The explanatory power is divided by the outcome-heterogeneity factor to get the impact factor – a quantitative measure of each case-mix adjusters' potential impact on the variance of the three global outcomes, standardized to the baseline variance in each of the three outcomes. If an outcome has very little baseline variance across the units of interest (facilities, in this case), an adjuster with a relatively low EP can potentially have a large impact on that outcome. Conversely, for outcomes that already exhibit substantial variance across the facilities, an adjuster would need relatively higher EP to have a noticeable impact.

presumably have more opportunities to observe what goes on in the nursing home. We could also argue that a respondent who visits more often may do so *because of* the quality of care delivered by the nursing home. For example, if the care were poor, the respondent might visit more often in order to supervise or supplement that care out of concern for the resident's health and wellbeing. Alternatively, if the care were very good, the respondent might visit more often because the experience is enjoyable. If so, frequency of visits would be endogenous with the quality scores and inappropriate as a case-mix adjuster. This possibility is enough of a concern that we propose rejecting this variable as an adjuster.

In sum, we propose that the remaining four variables be used as case-mix adjusters for the family survey: respondent age, respondent education, whether resident was permanently in the nursing home, and whether the resident was capable of making decisions.

Reporting to Nursing Homes

Based on the psychometric analyses and the case mix analysis, we developed reports with each nursing home presenting each nursing home's data and a comparison to the average. A sample report is in ***Appendix H***.

TEP review and finalization of instrument

After the analyzing the data and developing a proposed composite structure, AHRQ sent the draft composite structure to the TEP for review and held a conference call to obtain their input. The TEP had some additional questions and AIR provided input and additional analyses. Once the TEP reviewed it, the family survey was finalized and is presented in ***Appendix I, the Final Survey Instrument.***

References

Elliott, M.N., C. Edwards, J. Angeles, K. Hambarsoomians, & R.D. Hays, (2005). Patterns of unit and item nonresponse in the CAHPS Hospital survey. *Health Services Research, 40 (6), Part II:* 2096–2119.

Hurtado, M., Angeles, J., Blahut, S.A. & Hays, R., (2005). Assessment of the equivalence of the Spanish and English Versions of CAHPS® Hospital Survey. *Health Services Research, 40 (6) Part II:* 2140–2161.

Keller, S.D., O'Malley, A.J., Hays, R.D., Zaslavsky, A.M., Hepner, K.A. & Clearly, P.D. (2005). Methods used to streamline the CAHPS® Hospital Survey. *Health Services Research ,40(6) Part II*: 2057-2077.

O'Malley A.J., Zaslavsky, A.M., Hays, R.D. , Hepner, K.A., Keller, S.A. & Cleary, P.D.. (2005). Exploratory factor analyses of the CAHPS Hospital Pilot Survey responses across and within medical, surgical and obstetric Services. *Health Services Research*. *40(6), Part II*: 2078-2095.

O'Malley, A.J., Zaslavsky, A.M., Elliott, M.N., Zaborski, L. & Cleary. P.D. (2005). Case mix adjustment of the CAHPS Hospital Survey. *Health Services Research*. *40(6)*, *Part II*: 2162-2181.

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63,* 581-592.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons, Inc.

Tomarken, A.J. & Waller, N.G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. Annual Review of Clinical Psychology, (1), 31-65.

Zaslavsky, AM. (1998). Issues in case-mix adjustment of measures of the quality of health plans. *Proc Gov Soc Stat.* American Statistical Association, 56-64.

Measure #/Title/Steward

NH-027-10: Consumer Assessment of Health Providers and Systems (CAHPS®) Nursing Home Survey: Long-Stay Resident Instrument (Agency for Healthcare Research and Quality)

Description: The CAHPS® Nursing Home Survey: Long-Stay Resident Instrument is an in-person survey instrument to gather information on the experience of long stay (great 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

Initial In-Person Vote:

Recommended for time-limited endorsement with conditions – 10
Not recommended – 5
Abstained from vote - 5

Steering Committee Questions/Conditions for Measure Developer:	Response from Measure Developer
• In order to harmonize with the other nursing homes measures, the Steering Committee requested reconsideration of the definition of long-stay (over 100 days) and short stay (100 days or less) populations)	• The developer harmonized their definition of short and long stay residents with other NQF measures. Short-stay residents are defined as those within the home for 100 or fewer days and long term residents are those in the home for more than 100 days. • The denominator exclusions have been redefined in order to harmonize with the definitions used in other nursing home measures for short term residents (residents with a stay less than or equal to 100 days).
• Provision of cost information for long-term care resident surveys	• In general, when conducting large scale consumer satisfaction studies in long term care, there are economies of scale that can decrease the cost per interview. These include training local interviewers, thus saving on lodging, airfare, and payment for travel time; volume printing; capitalizing of the cost-effectiveness of scanning over data entry for large projects; and centralized and time limited project management. Vital Research has conducted 11 statewide long-term care resident surveys in three states (Minnesota, Ohio, and Rhode Island). Trained interviewers invited residents to participate in face-to-face interviews. Between 3,000 and 32,000 residents were interviewed during each statewide survey, depending on the state. The cost per completed interview ranged from $32 to $51. The cost per interview depended on the specific project requirements, such as the number of interviews (sampling plan), the number of nursing homes, distances between nursing homes (travel), and project deliverables (e.g. data analysis and reporting requirements). Each project also included fixed fees such as 3-day hands-on interviewer training sessions. Erin Pettegrew from the Ohio Department of Aging confirmed this Vital Research estimate and said that Ohio was at the lower range of costs (about $32). The state assesses $400 per year per nursing home to cover the costs of the survey and this is partially reimbursed by Medicaid.