A Dissertation

entitled

Assessment of the Measurement Properties of the NHCAHPS Family Survey:

A Rasch Scaling Approach

by

Matthew S. O'Connor

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Doctor of Philosophy Degree in Foundations of Education

_________________________________________
Christine M. Fox, PhD, Committee Chair

_________________________________________
Gregory E. Stone, PhD, Committee Member

_________________________________________
Noela Haughton, PhD, Committee Member

_________________________________________
John Gallick, PhD, Committee Member

_________________________________________
Patricia R. Komuniecki, PhD, Dean
College of Graduate Studies

The University of Toledo

May 2013
An Abstract of

Assessment of the Measurement Properties of the NHCAHPS Family Survey: A Rasch Scaling Approach

by

Matthew S. O'Connor

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Foundations of Education

The University of Toledo
May 2013

The introduction of the Consumer Assessment of Healthcare Providers and Systems (CAHPS), a family of survey instruments designed to capture and report people’s experiences obtaining health care could soon add satisfaction as a consistent dimension of quality that skilled nursing facilities (SNFs) are required to assess and report. The SNF setting has not yet been mandated to implement CAHPS for Nursing Homes (NHCAHPS). Given the critical implications (e.g., comparisons of SNFs on NHCAHPS scores) and decisions resulting from performance on the NHCAHPS survey (i.e., Medicare reimbursement) it was imperative to construct a measure of family satisfaction. The data were analyzed with the Rasch rating scale model. Rasch analyses demonstrate that the NHCAHPS Family Survey has adequate reliability, separation, fit, rating scale functioning, and dimensionality. Particular attention was given to the Overall Rating of Care item. According to the Rasch diagnostic indices, the NHCAHPS Family Survey composite domains did not function well. Results are discussed in terms of their application to usefulness by SNF management teams for process improvement and to include in the Five-Star Quality Rating System.
This dissertation is dedicated to my wife, Karen, and my children: Jason and Drew. Your tremendous patience and sacrifice as I pursued a lifelong dream will always be something I remember. I hope that one day Jason and Drew pick up this document and understand that when they have a dream they will need to work hard to accomplish it. Working hard will mean giving up immediate gratification to achieve something more fulfilling.

When I started on this doctoral journey, Jason was 5 and Karen was pregnant with Drew. This meant I had to leave Karen many evenings with 2 young boys as I attended class. She was able to make it through some tough times and will be happy this chapter of my life is now over. I know it was not easy and I will always be thankful.

There are many other people who were very helpful and supportive during this journey. My parents, Bill and Mary Pat, were always supportive of my dream. They helped me tremendously when this dream was derailed at one point in my life but knew it would be something I accomplished later. My in-laws, Ron and Helen Payeff, took the boys many evenings when I was gone and helped to Karen out more than I will ever know during this time. Thank you for your support.
Acknowledgements

First and foremost, I would like to thank my dissertation committee collectively for generously giving their time and feedback to me throughout this process. I would especially like to thank my chair, Christine Fox. Christine pushed me to develop my writing and research skills. I will always appreciate the amount of time she put in reviewing my drafts. Every time a draft was returned, I knew it would be filled with numerous “track changes” comments and grammatical changes that made my document better each time.

Next, I need to thank HCR ManorCare for their financial support to complete this degree. I will always remember Stephen Guillard as he made a significant company policy exception for me to achieve my dream. Also, I need to recognize John Huber as he provided me flexibility in my work schedule to attend courses that often began before the work day ended. Finally, John Gallick needs recognition for being a member of my committee and a support when I needed to discuss the project with someone. It is my desire to use what I have learned to continue to make HCR ManorCare the preeminent provider of long-term care.

I also need to thank Jamie Trabbic and Jennifer Orcelletto for putting up with me during the past year as I worked on this project. They listened to me vent about the process but were supportive of the goal often by picking up some extra work.
Table of Contents

Abstract

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Abbreviations

List of Symbols

I. Introduction

A. Five-Star Quality Rating System

B. Consumer Assessment of Healthcare Providers and Systems (CAHPS)

   a. NHCAHPS Family Survey Scoring and Reporting

C. Impact of Satisfaction Measurement

   a. Financial Implications

   b. Process Improvement

D. Instrument Development

E. Statement of the Problem

F. Purpose

G. Research Questions

II. The Current Methods for Collecting and Applying Quality Data in Skilled Nursing Facilities

A. Groups Advancing SNF Quality

B. Current View of Quality in Skilled Nursing Facilities
a. Annual Survey Inspection 19
b. Complaint Survey Investigations 20
c. Nursing Home Compare 22
d. Health Inspection Domain 22
e. Staffing Domain 26
f. Quality Measure Domain 29
g. Overall SNF rating (Composite Measure) 32

C. Issues with Current Quality Indicators 33
   a. Parsimony 34
   b. Measurement Issues 35
   c. Consumer-Determined Quality 36

D. The Rasch Model 39

E. Outcomes of Quality Indicators 40
   a. Pay for Performance 41
   b. Accountable Care Organizations 43

III. Methods and Procedures 45

A. Research Participants 45

B. Instrumentation 45

C. Data Management and Collection 47
   a. Phase 1: Facility Notification of the Survey 48
   b. Phase 2: Assignment of survey recipient 48
   c. Phase 3: Survey Notification and Distribution 48

D. Theoretical Framework for Measure Construction 49
a. Rating Scale Rasch Model 50

E. Data Analysis 52
   a. Reliability 52
   b. Validity 54

IV. Results 57
   A. Diagnostic Analyses 57
      a. Reliability Analyses 57
      b. Scale Category Analyses 58
         i. Rating scale type 1 58
         ii. Rating scale type 2 60
         iii. Rating scale type 3 62
         iv. Rating scale type 4 62
         v. Rating scale type 5 64
   B. Improving the Rating Scales 65
      a. Collapsing Recommendation to Others Scale 66
      b. Overall Rating of Care 69
   C. Diagnostic Analysis with Random Samples 71
      a. Scale Category Analysis 72
   D. Item Fit Analyses 73
   E. Variable Meaning 76
   F. Principle Contrasts Analysis 83
   G. Composite Domain Analysis 84
      a. Meeting Basic Needs Composite Domain 85
b. Nurses and Aides Kindness and Respect Towards Resident 85

c. Nursing Home Provides Information and Encourages Respondent Involvement 87

d. Nursing Home Staffing, Care of Belongings and Cleanliness 88

V. Discussion 89

A. Summary of Rasch Findings 90

B. Application of Findings 92

C. Improvement Recommendations 96

a. General Changes 96

b. Overall Rating of Care 97

D. Limitations 98

References 99

Appendices

A. NHCAHPS Family Survey 104

B. Composite Measure Proportional Scoring Method Description and Example 111

C. Composite Measures for the NHCAHPS Family Member Survey 113

D. Detailed Description of Scale Collapsing for Overall Quality of Care Item 115
List of Tables

Table 1  Five-Star Ratings Definitions. .................................................................4
Table 2  Types of NHCAHPS Family Survey Scores. .............................................8
Table 3  Health Inspection Score: Weights for Different Types of Deficiencies. ......24
Table 4  Weighting Factor for Annual Survey Inspections and Complaint Survey
        Investigations. ...............................................................................................25
Table 5  Weights for Repeat Revisits. .................................................................26
Table 6  National Average Hours per Patient Day Used in Calculation of Adjusted
        Staffing. ........................................................................................................27
Table 7  National Star Cut points for Staffing Measures. .....................................28
Table 8  Staffing Points and Five-Star Rating. .....................................................28
Table 9  MDS-Based Quality Measures. ...............................................................30
Table 10 Five Star Cut Points for MDS Quality Measure Domain Score. ..........31
Table 11 Five-Step Process for Assigning Overall Five-Star Rating. .................33
Table 12 Rating Scale Type 1 Diagnostics. ...........................................................58
Table 13 Rating Scale Type 2 Diagnostics. ...........................................................60
Table 14 Rating Scale Type 4 Diagnostics. ...........................................................62
Table 15 Rating Scale Type 5 Diagnostics. ...........................................................65
Table 16 Recommendation to Others 3-Point Rating Scale Diagnostics. ..........67
Table 17 Description of the Overall Rating of Care Collapsing Options. ..........70
Table 18 Summary of Changes in Person and Item Separation, Reliability,
        Variance Explained, and Fit Indices for Overall Care Rating Scale
        Collapsing Options, ..................................................................................71
Table 19  Separation, Reliability and Variance Explained for Random Samples. ..........72
Table 20  Range of the Item Fit Statistics for the Five Random Samples. ..................74
Table 21  Range of Item Measures and Standard Errors. ......................................77
Table 22  Standardized Residual Variance in Eigenvalue Units. .............................84
List of Figures

Figure 1  Sample webpage from the Nursing Home Compare website. ....................23
Figure 2  Rating Scale Type 1 Probability Curves. .............................................59
Figure 3  Rating Scale Type 2 Probability Curves. .............................................61
Figure 4  Rating Scale Type 4 Probability Curves. .............................................63
Figure 5  Rating Scale Type 5 Probability Curves. .............................................66
Figure 6  Recommendation to Others 3-Point Rating Scale Probability Curves. ....68
Figure 7  Person Map of Items. ........................................................................80
Figure 8  General Keyform Map. .....................................................................82
Figure 9  An Individual Survey Response for Service Recovery Intervention. ....94
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACA</td>
<td>Affordable Care Act</td>
</tr>
<tr>
<td>ACO</td>
<td>Accountable Care Organization</td>
</tr>
<tr>
<td>ADL</td>
<td>Activities of Daily Living</td>
</tr>
<tr>
<td>AHCA</td>
<td>American Health Care Association</td>
</tr>
<tr>
<td>AHRQ</td>
<td>Agency for Healthcare Research and Quality</td>
</tr>
<tr>
<td>AIR</td>
<td>American Institute of Research</td>
</tr>
<tr>
<td>CASPER</td>
<td>Certification and Survey Provider Enhanced Reports</td>
</tr>
<tr>
<td>CAHPS</td>
<td>Consumer Assessment of Healthcare Providers and Systems</td>
</tr>
<tr>
<td>CMS</td>
<td>Centers for Medicare and Medicaid Services</td>
</tr>
<tr>
<td>CSSR</td>
<td>Customer Service Survey Recipient</td>
</tr>
<tr>
<td>HCAHPS</td>
<td>Hospital Consumer Assessment of Healthcare Providers and Systems</td>
</tr>
<tr>
<td>HHCAHPS</td>
<td>Home Health Consumer Assessment of Healthcare Providers and Systems</td>
</tr>
<tr>
<td>IOM</td>
<td>Institute of Medicine</td>
</tr>
<tr>
<td>LPN</td>
<td>Licensed Practical Nurse</td>
</tr>
<tr>
<td>MDS</td>
<td>Minimum Data Set</td>
</tr>
<tr>
<td>MedPAC</td>
<td>Medicare Payment Advisory Commission</td>
</tr>
<tr>
<td>MIV</td>
<td>My InnerView</td>
</tr>
<tr>
<td>NHCAHPS</td>
<td>Nursing Home Consumer Assessment of Healthcare Providers and Systems</td>
</tr>
<tr>
<td>NRC</td>
<td>National research Corporation</td>
</tr>
<tr>
<td>NQF</td>
<td>National Quality Forum</td>
</tr>
<tr>
<td>PCC</td>
<td>Point Click care</td>
</tr>
<tr>
<td>P4P</td>
<td>Pay for Performance</td>
</tr>
<tr>
<td>QI</td>
<td>Quality Improvement</td>
</tr>
<tr>
<td>RN</td>
<td>Registered Nurse</td>
</tr>
<tr>
<td>RPCA</td>
<td>Rasch Principle Contrasts Analysis</td>
</tr>
<tr>
<td>RUG</td>
<td>Resource Utilization Group</td>
</tr>
<tr>
<td>SFF</td>
<td>Special Focus Facility</td>
</tr>
<tr>
<td>SNF</td>
<td>Skilled Nursing Facility</td>
</tr>
<tr>
<td>VBP</td>
<td>Value Based Purchasing</td>
</tr>
</tbody>
</table>
List of Symbols

\begin{itemize}
\item B_p \quad \text{Rasch Person Ability}
\item D_i \quad \text{Rasch Item Difficulty}
\item G \quad \text{Rasch Person and Item Separation Statistics}
\item S \quad \text{Observed Item Score}
\item P \quad \text{Proportion of Correct Items}
\end{itemize}
Chapter One

Introduction

More than 3 million elderly and disabled individuals will rely on services provided by a skilled nursing facility (SNF) at some point during the year, and among them 1.5 million will stay long enough to consider the SNF their main residence (Doshi, Shaffer, and Briesacher, 2005). These individuals, their families, and their friends count on SNFs to provide care that is of high quality. Enduring issues surrounding SNFs have been quality related. The often-poor quality of SNFs has been a consistent issue of concern for consumers, governments and researchers.

Numerous definitions of quality exist. The Institute of Medicine (IOM) (1996) provides one of the most widely cited definitions: “Quality is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge”. Definitions such as the one provided by the IOM offer guidance on what quality is, but operationalizing quality in SNFs can be problematic. The general nature and subjectivity of these definitions make the resulting quality indicators unable to fully realize the quality concept (Castle, Zinn, Brannon, and Mor, 1996). Also, most quality indicators have a medical or clinical focus that does not take into account the patient’s experience. Indicators of SNF quality continue to evolve with the purpose to utilize them across all SNFs.

Castle and Ferguson (2010) argue that, in many respects, there is no such thing as a typical SNF or a typical SNF patient. SNFs in the United States consist of a diverse group of providers. One of the most common differences in SNF providers is the type of patient subpopulations for whom they care. The most basic distinction between patient
subpopulations is long-stay patients and short-stay patients. Long-stay patients are often defined as those patients who remain at the facility for an extended period of time that is considered long enough to qualify the SNF as the main residence for the patient. Short-stay patients spend very little time (e.g., less than 30 days) in the facility. Short-stay patients are often at the facility for rehabilitation therapy.

SNFs appear to be moving towards a model that provides care for more short-stay patients. Over the past ten years, the percentage of patients who receive rehabilitation therapy care at a SNF has grown by more than 400% (Sangle, Buchanan, Cosenza, Bernard, Keller, Mitchell, Brown, Castle, Sekscenski, & Larwood, 2007). SNFs have both long-stay and short-stay patients residing in the facility. It may be more appropriate to think of the long-stay and short-stay distinction as a continuum (i.e., volume of short-stay patients) rather than a dichotomy.

There are financial incentives associated with caring for short-stay patients compared to traditional long-stay patients. For a traditional long-stay patient receiving Medicaid, the average daily reimbursement amount in 2010 was only $185. For a short-stay patient receiving Medicare, the average daily reimbursement rate is $560 per day and $416 for short-stay patients under a Managed Care plan.

**Five-Star Quality Rating System.**

A difference in short-stay patient volume is just one of many areas that make it difficult to define a “typical” SNF. Even though there is no typical SNF, the Centers for Medicare and Medicaid Services (CMS) enhanced its Nursing Home Compare website in 2008 to include a set of quality ratings to rate any SNF that participates in Medicare or Medicaid. These ratings take the form of several “star” ratings for each SNF. The
primary goal of this rating system is to provide patients and families with an easy way to understand assessment of SNF quality, making meaningful distinctions between high and low performing SNFs.

This rating system provides a graphical representation (i.e., stars) of overall high and low performance in three areas: Health Inspections, Staffing, and Quality Measures (CMS, 2010). In addition, the rating system features an overall five-star rating based on SNF performance in these three areas (see Table 1 for a formal definition of each area of the Five-Star Rating System (CMS, 2010)).

Patients and family members have had online access to the Five-Star Ratings as a source of information about the quality of the SNF for several years. This information is often used during the decision making process when selecting a SNF to receive care. In addition, several States (e.g., California) require SNFs to post their five-star ratings in the lobby area of the facility thus eliminating the need to access a SNF’s rating online.

As seen in the definitions of the each of the five-star ratings, patient and family member satisfaction is not used as an indicator of quality. Thus, the healthcare industry has begun to experience a philosophical change that now includes the patient and their experience as an integral part of quality. For example, the IOM puts the patient as central to the care system and assessment of quality (IOM, 2001). Also, collecting satisfaction information from SNF patients and family members may be important in improving some of the medical and clinical aspects of quality (Castle and Ferguson, 2010). The collection of satisfaction information has begun in with the introduction of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys in several area of the healthcare field (e.g., hospitals).
Table 1

*Five-Star Ratings Definitions*

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall SNF Rating</strong></td>
</tr>
<tr>
<td>The overall rating uses a formula that combines the five-star rating for</td>
</tr>
<tr>
<td>the Health Inspection, Staffing and Quality Measures.</td>
</tr>
<tr>
<td><strong>Health Inspections</strong></td>
</tr>
<tr>
<td>Health Inspections and based on the outcomes for State health inspections.</td>
</tr>
<tr>
<td>SNF ratings for the health inspections domain are based on the number,</td>
</tr>
<tr>
<td>scope, and severity of deficiencies identified during the three most</td>
</tr>
<tr>
<td>recent annual inspection surveys, as well as substantiated findings from</td>
</tr>
<tr>
<td>the most recent 36 months of complaint investigations. All deficiency</td>
</tr>
<tr>
<td>findings are weighted by scope and severity. This measure also takes</td>
</tr>
<tr>
<td>into account the number of revisits required to ensure that deficiencies</td>
</tr>
<tr>
<td>identified during the health inspection survey have been corrected.</td>
</tr>
<tr>
<td><strong>Staffing</strong></td>
</tr>
<tr>
<td>The Staffing rating is based on SNF staffing levels. These staffing levels</td>
</tr>
<tr>
<td>include RN hours per resident day and total staffing hours (RN + LPN +</td>
</tr>
<tr>
<td>nurse aide hours) per resident day. Other types of SNF staff such as</td>
</tr>
<tr>
<td>clerical, administrative, and housekeeping staff are not included in these</td>
</tr>
<tr>
<td>staffing numbers. These staffing measures are derived for the CMS</td>
</tr>
<tr>
<td>Certification and Survey Provider Enhanced Reports (CASPER) system.</td>
</tr>
<tr>
<td><strong>Quality Measures</strong></td>
</tr>
<tr>
<td>SNF ratings for the Quality Measures (QMs) are based on 9 of the 18 QMs</td>
</tr>
<tr>
<td>that are currently posted on the Nursing Home Compare website. These QMs</td>
</tr>
<tr>
<td>are based on Minimum Data Set (MDS) 3.0 resident assessments.</td>
</tr>
</tbody>
</table>

**Consumer Assessment of Healthcare Providers and Systems (CAHPS).**

The introduction of CAHPS, a family of survey instruments designed to capture and report people’s experiences obtaining health care, could soon add satisfaction as a consistent dimension of quality that SNFs are required to assess and report (Sangle, et. al, 2007). The CAHPS process has been introduced and federally mandated in several areas of healthcare, including hospital (HCAHPS) and home healthcare (HHCAHPS).
The SNF setting has not yet been mandated to implement CAHPS for Nursing Homes (NHCAHPS). The NHCAHPS surveys include two separate instruments for nursing home residents: one for those who live in a nursing home (Long-Stay Resident Survey) and another for those who have been discharged after a short stay (Discharged Resident Survey). The NHCAHPS surveys also include an instrument for gathering information on the experiences of the family members of residents in SNFs. The NHCAHPS Family Survey asks respondents to report on their own experiences (not the resident’s) with the SNF and their perceptions of the quality of care provided to a resident living in a SNF. The NHCAHPS Family Survey was developed to complement the Long-Stay Resident Survey. With a significant amount of influence on the placement decision of patient in a SNF falling to family members, SNF management teams may choose to collect data using the NHCAHPS Family Survey first (see Appendix A for a copy of the NHCAHPS Family Survey).

The development of these instruments was jointly supported by the Centers for Medicare and Medicaid Services (CMS) and the Agency for Healthcare Research and Quality (AHRQ). Like all CAHPS surveys, the instruments are in the public domain. In March 2011, the National Quality Forum (NQF) endorsed all three of these instruments as measures of nursing home quality, with the intention of mandating the use of NHCAHPS by 2014.

**NHCAHPS Family Survey scoring and reporting.** The survey items and scoring method for all CAHPS surveys have been imposed by CMS. The public reporting of scores uses composite domains. Composite domains combine results for
closely related items that have been grouped together. For example, the HCAHPS survey currently uses eight composite domains.

The calculation of the each composite domain score uses a proportional scoring method, which generates a proportion for each response option (see Appendix B for an example of the proportional scoring method). The top box proportion is then used as the composite domain score reported on the CMS website. Top box involves reporting only the composite proportion for the most positive response category of the items in the composite domain. For example, if a composite domain had three items with “Always” as the top box response, the proportion of respondents who answered “Always” to each item in the composite domain would be summed and divided by three to calculate the top box composite proportion.

The NHCAHPS Family Survey generates three types of satisfaction scores for reporting purposes (see Table 2 for a formal definition of each NHCAHPS score type). First, the Overall rating of Care is a single item on the survey that asks family members to assess the care at the SNF on a 0 to 10 scale. The second score type includes a set of four composite domains. The composite measures are: 1) Meeting Basic Needs: Help with Eating, Drinking and Toileting; 2) Nurses/Aides’ Kindness/Respect Towards Resident; 3) Nursing Home Provides Information/Encourages Respondent Involvement and Nursing Home Staff; and 4) Care of Belongings and Cleanliness (see Appendix C for a list of the survey items included in each composite measure). The third score type includes additional survey items that do not fit into one of the composite domains.
Impact of Satisfaction Measurement.

If the NHCAHPS Family Survey contains poorly constructed items, and family member ratings are summarized by an average of these items, as described in the scoring of the four composite domains, then there is no basis for knowing what is being measured and no basis for comparing the results between SNFs. The results obtained from the NHCAHPS survey have several high stakes implications (i.e., Financial and Process Improvement) that call for the need to examine the validity of using these items in this way to construct a measure of family satisfaction used in high-stakes decision-making.

Financial implications. Results from the measurement of patient and family member satisfaction could have a direct financial impact on the SNF. Patient/family member satisfaction is an important indicator of a healthcare provider’s growth and profitability (Milutinovic, Brestovacki, Martinov-Cvejin, 2009). The direct financial implications satisfaction measurement has already begun in the hospital setting. Beginning in 2012, CMS started to withhold 1% of Medicare payments to hospitals partially based on scores from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) Survey. HCAHPS is a federally mandated survey instrument designed to capture and report people’s experiences obtaining health care in hospitals. Hospitals would have this withholding returned by achieving HCAHPS composite domain scores above the national average, whereas those hospitals that do not achieve composite scores above the national average would not have their 1% withholding returned to them. CMS plans to increase this withholding by .25% each year for the next six years.
Table 2

*Types of NHCAHPS Family Survey Scores*

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall Rating of Care</strong></td>
</tr>
<tr>
<td>This score is based on the “Overall Satisfaction” survey item that uses a</td>
</tr>
<tr>
<td>scale of 0 to 10 to measure family members’ assessment of the SNF. This</td>
</tr>
<tr>
<td>score is sometimes referred to as the “global rating” or “overall rating”</td>
</tr>
<tr>
<td><strong>Composite Domains</strong></td>
</tr>
<tr>
<td>These are also known as rating composites. Composite domains combine</td>
</tr>
<tr>
<td>results for closely related items that have been grouped together.</td>
</tr>
<tr>
<td>Composite domains are strongly recommended for both public and private</td>
</tr>
<tr>
<td>reporting because they keep the reports comprehensive yet of reasonable</td>
</tr>
<tr>
<td>length. There are four composite domains for the Family Survey:</td>
</tr>
<tr>
<td>• Meeting Basic Needs: Help with Eating, Drinking and Toileting</td>
</tr>
<tr>
<td>• Nurses/Aides’ Kindness/Respect Towards Resident</td>
</tr>
<tr>
<td>• Nursing Home Provides Information/Encourages Respondent Involvement</td>
</tr>
<tr>
<td>• Nursing Home Staff, Care of Belongings and Cleanliness</td>
</tr>
<tr>
<td>The calculation of each composite uses a proportional scoring method,</td>
</tr>
<tr>
<td>which generates a proportion for each response option. This approach</td>
</tr>
<tr>
<td>allows two options for reporting, average score and top box. Top box</td>
</tr>
<tr>
<td>involves reporting only the score for the most positive response category.</td>
</tr>
<tr>
<td><strong>Individual Items</strong></td>
</tr>
<tr>
<td>These are survey items that did not fit into one of the composite domains.</td>
</tr>
<tr>
<td>Scores from these items may be included in public reports, but they are</td>
</tr>
<tr>
<td>especially useful in reports for individual facilities and other internal</td>
</tr>
<tr>
<td>audiences that use the data to identify specific strengths and weaknesses.</td>
</tr>
</tbody>
</table>

CMS has indicated that a similar withholding would occur when SNFs are      |
federally required to collect and report satisfaction information. In addition to potential Medicare withholding, some individual states have already begun to penalize SNFs by decreasing Medicaid reimbursement rates through individual State run satisfaction survey
processes. Many of these states have indicated that they will eliminate their State level satisfaction survey processes and replace it with results from a federally mandated survey to adjust Medicaid rates. Therefore, the financial impact would be particularly strong for SNFs because Medicare and Medicaid are their two primary sources of revenue.

Results from the measurement of family member satisfaction could also have an indirect financial impact. The indirect financial impact would result from the public reporting of CAHPS composite domains. For example, each hospital’s HCAHPS composite domain results are available on the website www.hospitalcompare.gov. Patients are encouraged to research this site as another source of information when determining which hospital they are going to receive a surgical or medical procedure. When a hospital’s HCAHPS scores are lower than other hospitals in the area, it could impact a patient’s decision thus resulting in lost business through the patient’s choice of a competitor.

SNF satisfaction survey results will be posted on the CMS Nursing Home Compare website, with the intent to provide similar information to patients and family members when choosing a nursing home. These scores could be reported separately from the five-star ratings or be integrated into the Five-Star Rating System. The public reporting composite domain scores would be particularly important to family members because family members are frequently the decision makers when it comes to placing someone in a SNF (Schulz, Belle, Czaja, McGinnis, Stevens, & Zhang, 2004). This means that a SNF could lose business when family members review the satisfaction scores reported on this site, which further reinforces the need to develop a measure of family satisfaction.
**Process improvement.** Results from the NHCAHPS Family Survey are used by SNF management teams to identify and improve factors that are important to the family member’s experience at the SNF. This process of using the results found in the feedback report is an important part of a SNF’s quality improvement activities.

By understanding the factors that are important to family members’ experience in SNFs, a facility can use the information found in the feedback report to improve its reputation in the market. With an improved reputation, a SNF may get more family members to consider it when looking to place someone in a SNF.

SNF management teams spend time reviewing feedback reports from satisfaction surveys they conduct. The feedback report provides information on the individual survey items with the intent of using this information to develop improvement plans. Feedback reports typically provide SNF management teams with two primary sources of information to use when developing improvement plans. The first source of information is the *item-score* table that rank orders the survey items from the highest scoring to the lowest scoring. The second source of information is the *key drivers* table. The *key drivers* table rank orders the survey items based on their correlation with the Overall rating of Care. It is stated in the feedback report that improving performance on items located at the top of the *key drivers* list will increase scores on the Overall rating of Care.

DerGurhian (2009) reported that the current practice employed by many SNF teams involves the examination of items with the highest correlation with overall satisfaction and the subsequent development of improvement plans in those areas identified by the correlations. Without the development and implementation of valid measures of family member satisfaction, SNF management teams may be making
decisions on areas for improvement based on statistical analyses while disregarding the underlying meaning of the ratings (what family members were trying to communicate about their level of satisfaction), further undermining management’s understanding of what it means to move towards continuous improvement.

Given the financial and process improvement implications of scores resulting from the NHCAHPS Family Survey, it is important to understand the how the instrument was developed. SNF management teams and family members who are using this information in the selection of a SNF may assume that the statistical inferences being made from the NHCAHPS Family Survey are both reliable and valid.

**Instrument Development.**

The development of the Family Survey followed the standard CAHPS process by conducting a literature review and focus groups, performing cognitive testing of draft survey items, obtaining stakeholder input, conducting a field test of a draft instrument and survey administration protocol, and undertaking psychometric analyses of field test data. Stakeholder input was obtained by establishing a Technical Expert Panel composed of industry, regulators and quality improvement organizations, payers, long-term care researchers, and consumer advocates (cite from the CAHPS Nursing Home Survey and Instructions Manual). The NHCAHPS Development Team encouraged researchers to conduct additional testing and development of the survey. This additional testing should include an assessment of the extent to which the NHCAHPS Family Survey can be used to construct a linear measure of family satisfaction.

Published NHCAHPS research has focused on the resident satisfaction instrument and not on the family satisfaction instrument (Sangle et. al, 2007). The only published
research using the NHCAHPS Family Survey instrument comes from the American Institute of Research (AIR) Final Report (2008) that was presented to Agency for Healthcare Research and Quality (AHRQ). This report focused on the development of the four composite domains presented in Table 2.

In the AIR Final Reports, the researchers indicated that the four proposed domains included: a balance of theory (original intent of items and composites), statistical evidence of reliability and validity (item-level and nursing home level, factor analyses) and stakeholder perspectives. The internal consistency reliability estimates for composite domains ranged from 0.73 to 0.81. These estimates led the AHRQ to endorse the NHCAHPS Family Survey as a measure of family satisfaction and indicated that it is both reliable and valid for its intended purpose, resulting in confident use by SNF organizations as a way to assess family satisfaction.

**Statement of the Problem.**

The approaches used for the NHCAHPS instrument development are methodologically flawed. Due to the impact results from the NHCAHPS Family Survey will have on a SNF’s growth and identification of areas for family satisfaction improvement, it is important to evaluate the extent to which data from the survey question produces information that is meaningful for making decisions. The NHCAHPS Family Survey development is based on assumptions about the types of items and the ways in which they should be numerically combined. Before this survey is implemented across all SNFs these assumptions need to be empirically tested.

In addition, the Overall rating of Care item is often given the most attention by family members as a summary of the care provided. With the transparency of these
scores increasing (i.e., presented on public websites) there is a danger in simply providing a potentially unfair assessment of family members’ perception of care with uncorrected raw scores. The Overall rating of Care item and the composite measure scores are used to assign ratings to SNFs in high stakes settings where assessor ratings (i.e., family member survey responses) can result in revenue loss or lawsuits due to the inherent error in human judgment.

Also, SNF teams rely on NHCAHPS Family Survey feedback reports to provide insight into the areas that impact the Overall rating of Care item. Feedback reports identify the key drivers with the implication that improving these areas will lead to higher scores on the Overall rating of Care item. However, the statistical analyses used to identify such items suffer from two major problems. First, researchers perform mathematical operations (e.g., means and correlations) on ordinal-level (rating scale) data collected by the patient satisfaction survey. This statistical approach, coupled with the use of small samples (e.g., when facility level analyses are conducted), produces results that are based on small fluctuations or idiosyncratic responses to the survey. Decisions regarding which items to focus on are then made based upon inaccurate empirical guidelines. Second, decision-making based purely upon statistical analyses disregards the underlying meaning of the ratings (what family members were trying to communicate about their level of satisfaction), further undermining management’s understanding of what it means to move towards continuous improvement.

These problems call for a different analytic approach to analyzing rating scale data while also taking into account the pattern of responses to better understand the meaning of family member satisfaction. Such an approach will allow for decisions to be
made based on empirically justifiable statistics (quantitative evidence), contextualized within an understanding of the construct of patient satisfaction (qualitative evidence). This approach will reduce the probability that a SNF that provides high quality care will receive low ratings due to unlucky encounters with severe assessors (i.e., family members) or that SNFs that provide low quality care will receive high ratings due to lucky encounters with assessors.

Given the critical implications (i.e., comparisons of facilities NHCAHPS scores) and decisions resulting from performance on the NHCAHPS Family Survey (i.e., Medicare reimbursement) it is imperative that steps are taken to determine the extent to which the NHCAHPS Family Survey is a measure of family satisfaction. The information currently collected with the NHCAHPS Family Survey are simply descriptive numbers (e.g., 20% agree with an item) or are correlated with one another to examine response patterns. These descriptions and correlations are not measures and hence need to be converted into measures. In other words, the NHCAHPS Family Survey responses need to be examined with a scientific measurement model to construct meaningful and replicable numbers.

**Purpose.**

An effective measure of family member satisfaction provides SNF management teams with an understanding of the relationship between persons and items that operationalizes family satisfaction when developing improvement plans from satisfaction survey feedback reports. Improving SNF family member satisfaction has important financial and growth implications for a facility.
The purpose of this study is to use the NHCAHPS Family Survey as a beginning point in the development of a measure of family member satisfaction. This approach will provide insight into potential changes required for the NHCAHPS Family Survey to provide meaningful inferences from patterns of responses at the construct level; family satisfaction.

**Research Questions.**

1) To what extent do the respondents use the NHCAHPS Family Survey rating scale categories as intended?

2) What rating scale structure is implied by the respondents’ use of Overall rating of Care?

3) How well do the items separate respondents into statistically distinct and meaningful categories?

4) To what extent do the items form a reliable (stable) line of inquiry (ruler)?

5) To what extent is the NHCAHPS Family Survey measuring a unidimensional construct?

6) Is the item ordering meaningful?

7) To what extent are the composite measures assessing unidimensional constructs?
Chapter Two
The Current Methods for Collecting and Applying Quality Data in Skilled Nursing Facilities

The Institute of Medicine (IOM) report (2001) entitled “Crossing the Quality Chasm: A New Health Care System for the 21st Century” significantly raised the bar in terms of the level of quality that should be expected from the entire health care system in the United States. The conceptualization of quality in the IOM reports was particularly broad, encompassing patient safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity as the six “Aims of Quality.”

Berwick (2009) offered two interpretations of the Aims of the IOM report. First, Berwick (2009) offered a “technocratic” interpretation in which only two of the six IOM Aims, safety (avoiding harm) and effectiveness (avoiding medicine overuse and underuse), are primary, whereas the others are important only to the extent to which they are related to safety and effectiveness. Second, Berwick (2009) offered a “consumerist” interpretation, which takes each of the six Aims on its own merits. In a “consumerist” approach, the patient, rather than the staff or a third-party governing party (e.g., CMS), determines whether quality standards are being met at the health care facility.

Over the last five years, many areas of health care report that patient safety has assumed primacy over the other Aims (Schwartz, Cramer, Holmes, Cohen, Restuccia, Lukas, Sullivan & Charns, 2010). Safety, defined as not causing harm, should undoubtedly be the minimum a patient should expect from the health care system. The importance attached to this Aim reflects how far the health care system is from providing a safe environment in many areas of health care. However, because of the importance of
the other Aims noted in the IOM report, at some point health care is likely to gravitate toward the consumerist-based conceptualization of quality (Schwartz et al., 2010).

One area of health care that has seen a continual movement toward a “consumerist” definition of quality is long-term care. A “consumerist” approach to quality is needed in long-term care due to the increasing demand of services provided by skilled nursing facilities (SNFs). This increase in demand for services has led consumers (i.e., patients and families) and several other groups to focus on improvement in all six Aims in the IOM report.

**Groups Advancing SNF Quality.**

There are four primary audiences working to improve quality measurement and performance in long-term care: providers themselves, regulators, purchasers, and consumers. The value and purpose of increased SNF quality varies as a function of these audiences (Frankenfield, Marciniak, Drass, and Jencks, 1997).

SNF providers have consistently been the strongest advocates for finding ways to measure quality in order to identify care problems as part of their continuous quality improvement (QI) program (Mor, Berg, Angelelli, Gifford, Morris, and Moore, 2003). The establishment of quality indicators based on uniformly available data has been the basis for QI programs in SNFs for many years (Castle, 1999).

SNF providers are also pushing for increased quality because SNF patients are arriving with greater medical complexity and require more extensive and costly care (Brieracher, Field, Baril, and Gurwitz, 2009). To appropriately care for a higher-acuity, short-stay patient population and a more frail and unstable long-stay resident population,
SNF operators are taking steps to improve the delivery of the clinical and hospitality services (Cantlupe, 2012).

The increase in the volume and complexity of patients has led to increased demand from government regulatory agencies (e.g., The Centers for Medicare & Medicaid Services (CMS)) to quickly advance how quality is defined in SNFs and the efforts by SNF operators to improve quality at their facilities. CMS provides quality information to guide the facility survey and certification process that is accomplished by state departments of health throughout the country. Reports of a SNF’s performance on numerous dimensions of quality are provided to the regulatory inspectors to guide the inspection process to focus on identified quality problems (Mor, et al., 2003).

Purchasers of health care (e.g., insurance companies) often urge SNF providers to compete on both price and quality. This competition on quality has also led to the development and implementation of pay-for-performance programs that are designed to reward SNFs for achieving high levels of quality performance or improvements in quality.

Finally, patients, their families, and advocates have periodically called for the public release of quality data so they can actively select the SNF providers that best meet their needs. More than 88% of all SNF residents are older than 65, and 45% are 85 or older, so the rate of growth in this age group affects demand for SNF care. IBISWorld (2011) estimated that the number of adults aged 65 or older would grow at an average annual rate of 2.2% during the five years to 2011 and reach 41.5 million. The U.S. population is aging, resulting in increased demand for SNF care because this demographic is more prone to injuries and illnesses that require assistance with activities.
of daily living (ADLs) (IBISWorld Industry Report, 2011). This increase in demand has led to continued development and reporting of quality in SNFs that is currently used by patients and families to compare SNF performance.

**Current View of Quality in Skilled Nursing Facilities.**

The quality of SNFs is generally assessed using several quality indicators. These current sets of quality indicators are important because they are national in scale and include a fairly comprehensive scope of quality indicators. In addition, these quality indicators influence which areas SNF providers address, regulators examine, and patients scrutinize. Two of the most widely used quality indicators are deficiency citations assessed during annual survey inspections and complaint survey investigations. Deficiency citations are also included with other additional quality indicators reported on the Nursing Home Compare website (Castle, et al., 2010).

**Annual survey inspection.** CMS and each state’s health department visit SNFs on a regular basis (e.g., annually) to assess the quality of care that Medicare and Medicaid requires each SNF to provide. Survey teams spend several days in a SNF to identify deficiencies in the quality of care that is provided. The areas that are assessed include medication management, proper skin care, assessment of resident needs, SNF administration, environment, food services, resident rights, quality of life, and any deficiencies in meeting CMS safety requirements (such as protection from fire hazards). When deficiencies are identified, CMS requires each problem to be corrected. If serious problems are not corrected, CMS may terminate the SNF’s participation in Medicare and Medicaid.
Deficiency citations are influential quality indicators because they represent an assessment of quality coming from the main SNF oversight body (Castle, et al., 2010). Deficiency citations are also presented in many report cards (e.g., Nursing Home Compare), in government reports, and in the lay press (e.g., daily newspapers).

Most SNFs have some deficiencies, with the average being 6-7 deficiencies per survey (CMS Five-Star Users Guide, 2012). Problems are typically corrected within a reasonable period of time. However, CMS has identified a minority of SNFs that do not meet the typical profile for number and severity of deficiencies. These SNFs often have twice the average number of deficiencies, more serious deficiencies (including harm and injury to patients), and a pattern of serious problems that has persisted over a period of at least 3 years (CMS Five-Star Users Guide, 2012).

Although such SNFs would periodically institute enough improvements in the presenting problems that they would be in substantial compliance on one annual survey, significant problems would often resurface by the time of the next annual survey. Such facilities with an “in and out” compliance history rarely addressed underlying systemic problems that result in repeated cycles of serious deficiencies. To address this problem, CMS created the “Special Focus Facility” (SFF) initiative.

CMS requires that SFF nursing homes be visited in person by survey teams twice as frequently (about twice per year). The longer the problems persist, the more stringent CMS enforcement actions that are employed. Examples of such enforcement actions are civil monetary penalties (“fines”) or termination from Medicare and Medicaid. In addition, consumers have access to a list of SNFs identified as SFF facilities.

**Complaint survey investigations.** In addition to the annual state health department visits, consumers (i.e., patients and families) can call the department of health
to initiate a complaint investigation. Monitoring and investigating complaints about SNF care serves as an important supplemental role to the annual inspections required of all Medicare and Medicaid providers. By definition, complaint investigations infuse the perspectives of SNF patients and their families into the formal oversight process and, unlike annual surveys, can occur anytime (Stevenson, 2006). The number of complaint investigations between annual survey inspections provides a timely signal of problems that could arise in subsequent annual inspections.

Complaint survey investigations appear generally useful in assessing SNF quality, but there are challenges that arise when trying to compare individual SNFs and when trying to distinguish between low-complaint SNFs. The national average of 4.2 complaints per 100 residents per year implies that only around four complaint survey investigations occur per SNF annually (Stevenson, 2006). The median rate is even lower at 3.2 complaint survey investigations per year. Low complaint survey investigation rates are often expected because complaint surveys are self-reports and require action by patients or family members to express their dissatisfaction.

A SNF with no complaint survey investigations could be viewed as a sign of quality care or as a sign of problematically low rates of complaints due to fear of retaliation by the staff, despite the fact complaints can be given anonymously (Zimmerman, Hawes, Stegeman, and Bowers, 2003). Low complaint survey investigation rates result in SNFs looking identical in their complaint profiles. In addition, complaints are, by definition, a negative measure, and their absence does not necessarily imply high-quality SNF care. Although quality SNF care has been characterized by the absence of negative events, recent efforts have been made to assess
and monitor positive outcomes for quality of care for SNF patients (Mor, et al., 2003). Complaint survey investigations offer regulators and purchasers of health care an additional low-cost means to monitor SNF quality and provide consumers salient information to do the same.

Deficiency citations and complaint survey investigations are often examined independently as indicators of SNF quality, especially by consumers looking to identify a place to receive care. CMS has also taken steps to combine this information with other important quality information through the introduction of its Nursing Home Compare reporting site.

**Nursing home compare.** Nursing Home Compare was developed by CMS to provide information on all Medicare- and Medicaid-certified SNFs in the United States. The information provided on the Nursing Home Compare website takes the form of several “star” ratings for each SNF. This rating system provides a graphical representation of overall high and low performance in three areas: Health Inspections, Staffing, and Quality Measures (CMS, 2010). The intent of the Five-Star Quality Rating System is to provide valuable and comprehensible information to consumers based on the best data currently available. An example of the rating information included on Nursing Home Compare is shown in Figure 1. Users of the website can obtain additional information on the SNF’s performance within each domain.

**Health inspection domain.** The Health Inspection Domain uses information from the deficiency citations and complaint survey investigations described above. The number of stars a SNF receives for the Health Inspection Domain is calculated based on points assigned to deficiencies identified in each SNF’s current annual survey inspection
and the two prior annual surveys, as well as substantiated deficiency findings from the most recent three years of complaint survey investigations (CMS, 2012). In addition, the number of revisits required to confirm that correction of deficiencies has restored compliance is also used to calculate the number of stars a SNF receives in the Health Inspection Domain.

Figure 1. Sample webpage from the Nursing Home Compare website.

Points are assigned to individual health deficiencies according to their scope and severity. More points are assigned for deficiencies that are deemed serious and widespread, whereas fewer points are assigned for less serious and isolated deficiencies.
(See Table 3). If the deficiency generates a finding of substandard quality of care, additional points are assigned.

Table 3

*Health Inspection Score: Weights for Different Types of Deficiencies*

<table>
<thead>
<tr>
<th>Severity</th>
<th>Isolated</th>
<th>Pattern</th>
<th>Widespread</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate jeopardy to resident health or safety</td>
<td>J</td>
<td>K</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>50 points</td>
<td>100 points</td>
<td>150 points</td>
</tr>
<tr>
<td></td>
<td>(75 points)</td>
<td>(125 points)</td>
<td>(175 points)</td>
</tr>
<tr>
<td>Actual harm that is not immediate jeopardy</td>
<td>G</td>
<td>H</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>20 points</td>
<td>35 points</td>
<td>45 points</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(40 points)</td>
<td></td>
</tr>
<tr>
<td>No actual harm with potential for more than minimal harm that is not</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>immediate jeopardy</td>
<td>4 points</td>
<td>8 points</td>
<td>16 points</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(20 points)</td>
</tr>
<tr>
<td>No actual harm with potential for minimal harm</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>0 point</td>
<td>0 point</td>
<td>0 point</td>
</tr>
</tbody>
</table>

Note: Figures in parentheses indicate points for deficiencies that are substandard quality of care.

Source: Centers for Medicare & Medicaid Services

The calculation used to determine the Health Inspection Domain score uses a weighting formula that weighs recent survey results more heavily than earlier surveys.

Table 4 provides a summary of the weighting factors used for annual survey inspections and complaint survey investigations.
There are some surveys that appear in both the annual inspection and the complaint investigations. To avoid the potential double-counting, deficiencies that appear on the complaint surveys conducted within 15 days before or after an annual survey are counted only once. If the scope or severity differs on the two surveys, the highest scope-severity combination is used.

After the points and weights of deficiencies are determined, the second part of the Health Inspection Domain score is determined by the number of revisits required to confirm correction of the deficiencies. No points are assigned for the first revisit. Points are assigned for the second, third, and fourth revisits and are proportional to the Health Inspection Score (See Table 5).

If a SNF fails to correct deficiencies by the time of the first revisit, these additional revisit points are assigned up to 85% of the total health inspection score. CMS’ experience is that SNFs that fail to demonstrate restored compliance with safety and quality of care requirements during the first revisit have lower quality of care than other SNFs (Castle and Ferguson, 2010). In other words, additional revisits are often associated with more serious quality problems.

Table 4

Weighting Factor for Annual Survey Inspections and Complaint Survey Investigations

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Weight</th>
<th>Timeframe</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Recent</td>
<td>½</td>
<td>Most recent 12 Months</td>
<td>½</td>
</tr>
<tr>
<td>Previous Year</td>
<td>1/3</td>
<td>13-24 Months ago</td>
<td>1/3</td>
</tr>
<tr>
<td>Prior Year</td>
<td>1/6</td>
<td>25-36 Months ago</td>
<td>1/6</td>
</tr>
</tbody>
</table>
Table 5

Weights for Repeat Revisits

<table>
<thead>
<tr>
<th>Revisit Number</th>
<th>Noncompliance Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>0</td>
</tr>
<tr>
<td>Second</td>
<td>50 percent of the Health Inspection Score</td>
</tr>
<tr>
<td>Third</td>
<td>70 percent of the Health Inspection Score</td>
</tr>
<tr>
<td>Fourth</td>
<td>85 percent of the Health Inspection Score</td>
</tr>
</tbody>
</table>

**Staffing domain.** There is considerable evidence between SNF staffing levels, staffing stability, and patient health outcomes (Stevenson, 2006). Kramer and Fish (2001) identified ratios of staff to patients below which patients are at substantially higher risk of quality problems. Findings such as those reported above have resulted in the development of a staffing domain reported on the Nursing Home Compare site to allow patients to identify SNFs with higher levels of staffing.

The source data for the Staffing Domain is derived from the CMS Certification and Survey Provider Enhanced Reports (CASPER) system. The rating for the Staffing Domain is based on two case mix-adjusted measures (CMS, 2012). First, the total nursing hours per patient day are used. Total nursing hours include RN, LPN, and nurse aide hours. Second, RN hours per resident day are also used independent of the total nursing hours as a rating for the staffing domain. RN hours include registered nurses, the RN director of nursing, and nurses with administrative duties at the SNF.

As noted above, the total nursing hours and RN hours are adjusted based on the Resource Utilization Group (RUG-III) case-mix system. The following formula is used to calculate the case-mix adjusted hours per patient day for each SNF:
Hours Adjusted = (Hours Reported + Hours Expected) * Hours National Average

In the above formula, Hours National Average is the mean across all SNFs of the reported hours per patient day for a given staff type. Table 6 provides Hours National Average for total nursing hours and RN hours per patient day. Hours Expected is based on the distribution of patients by RUG-III group in the given quarter closest to the date of the most recent annual survey inspection. Hours Reported is provided on CMS form 671.

Table 6

*National Average Hours per Patient Day Used in Calculation of Adjusted Staffing*

<table>
<thead>
<tr>
<th>Type of Staff</th>
<th>National Average Hours Per Patient Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Nursing Hours (RNs + LPNs + Nurse Aides)</td>
<td>4.0309</td>
</tr>
<tr>
<td>Registered Nurses (RNs) Hours</td>
<td>0.7472</td>
</tr>
</tbody>
</table>

Source: Centers for Medicare & Medicaid Services

The two staffing measures are given equal weight. A rating of 1 to 5 stars is assigned for the total nursing hours and the RN hours based on a percentile-based method. CMS identified the percentile cut points using staffing data from December 2011. New cut points will be calculated every two years. Table 7 provides the cut points used for the two staffing measures.

The overall staffing Five-Star rating is based on a combination of the RN and total nurse staffing ratings shown in Table 7. To receive a Five-Star rating, SNFs must meet
or exceed the Five-Star levels noted in the last column of Table 7. Table 8 shows the
total staffing Five-Star rating based on a combination of the two staffing ratings.

Table 7

National Star Cut Points for Staffing Measures

<table>
<thead>
<tr>
<th>Staff Type</th>
<th>1 Star</th>
<th>2 Stars</th>
<th>3 Stars</th>
<th>4 Stars</th>
<th>5 Stars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower</td>
<td>upper</td>
<td>lower</td>
<td>upper</td>
<td>lower</td>
</tr>
<tr>
<td>RN</td>
<td>0.283</td>
<td>≥ 0.283</td>
<td>&lt; 0.379</td>
<td>≥ 0.379</td>
<td>&lt; 0.513</td>
</tr>
</tbody>
</table>

Source: Centers for Medicare & Medicaid Services

Table 8

Staffing Points and Five-Star Rating

<table>
<thead>
<tr>
<th>RN Hours</th>
<th>Total Staffing</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.283</td>
<td>1-star</td>
</tr>
<tr>
<td>0.282-0.378</td>
<td>1-star</td>
</tr>
<tr>
<td>0.379 – 0.512</td>
<td>2-stars</td>
</tr>
<tr>
<td>0.513 – 0.709</td>
<td>2-stars</td>
</tr>
<tr>
<td>≥ 0.710</td>
<td>3-stars</td>
</tr>
</tbody>
</table>

Source: Centers for Medicare & Medicaid Services
**Quality Measure Domain.** A set of quality measures has been developed for Minimum Data Set (MDS)-based indicators to describe the quality of care being provided at the SNF. The SNF rating on the Quality Measure Domain is based on performance on a subset of 9 out of a possible 18 quality measures collected in the MDS. The measures were selected based on their validity, reliability, the extent to which the measure is under the SNF’s control, statistical performance, and importance (CMS, 2012). Seven of the 9 quality measures are related to long-stay patients, and the remaining 2 measures are related to short-stay patients. See Table 9 for a full definition of each of the 9 measures.

Ratings for the Quality Measure Domain are calculated using the three most recent quarters for which data are available. The use data from three quarters is intended to increase the number of assessments available for calculating the quality measure rating, increase the stability of the estimates, and reduce the amount of missing data (CMS, 2012).

For each of the long-stay and short-stay measures reported in Table 7, 1 to 100 points are assigned based on the SNF’s performance. SNFs achieving the best possible score on a quality measure (i.e., 0% of patients triggering the quality measure) are assigned 100 points. The remaining SNFs are assigned 1 to 99 points, based on national percentiles of the quality measure distribution for SNFs with values greater than 0%. SNFs that score in the top 1% (of those with non-zero values) score 99 points. SNFs in the lowest 1% receive 1 point. All of the 9 quality measures are given equal weight. The points are summed across all quality measures to create a total score for the SNF. The total possible score ranges from 9 to 900 points.
<table>
<thead>
<tr>
<th>Long-Stay Measure</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of patients whose need for help with daily activities has increased</td>
<td>This measure reports the percent of long-stay patients whose need for help with late-loss activities of daily living (ADLs) has increased when compared to the prior assessment. This is a change measure that reflects worsening performance on at least 2 late-loss ADLs by one functional level or on one late-loss ADL by more than one functional level compared to the prior assessment. Late-loss ADLs include bed mobility, transfer, eating, and toileting.</td>
</tr>
<tr>
<td>Percent of high-risk patients with pressure ulcers</td>
<td>This measure captures the percentage of long-stay, high-risk patients with Stage II-IV pressure ulcers. High-risk patients for pressure sores are those who are impaired in bed mobility or transfer, who are comatose, or who suffer from malnutrition.</td>
</tr>
<tr>
<td>Percent of patients who have/had a catheter inserted and left in their bladder</td>
<td>This measure reports the percentage of patients who have had an indwelling catheter in the last 7 days. Indwelling catheter use may result in complications, such as urinary tract or blood infections, physical injury, skin problems, bladder stones, or blood in the urine.</td>
</tr>
<tr>
<td>Percent of patients who were physically restrained</td>
<td>This measure reports the percent of long-stay nursing facility patients who are physically restrained on a daily basis. A patient who is restrained daily can become weak, lose his or her ability to go to the bathroom without help, and develop pressure sores or other medical complications.</td>
</tr>
<tr>
<td>Percent of patients with a urinary tract infection</td>
<td>This measure reports the percent of long-stay nursing facility patients who have had a urinary tract infection within the past 30 days.</td>
</tr>
<tr>
<td>Percent of patients who self-report moderate to severe pain</td>
<td>This measure captures the percent of long-stay patients who report either (1) almost constant or frequent moderate to severe pain in the last 5 days or (2) any very severe/horrible pain in the last 5 days.</td>
</tr>
</tbody>
</table>
Percent of patients experiencing one or more falls with major injury

This measure reports the percent of patients who experience one or more falls with major injury (e.g., bone fractures, joint dislocations, closed head injuries with altered consciousness, or subdural hematoma) in the last year (12-month period).

**Short-Stay Measures**

Percent of patients with pressure ulcers that are new or worsened

This measure captures the percentage of short-stay patients with new or worsening Stage II-IV pressure ulcers.

Percent of patients who self-report moderate to severe pain

This measure captures the percent of short-stay patients with at least one episode of moderate/severe pain or horrible/excruciating pain of any frequency in the last 5 days.

Source: Centers for Medicare & Medicaid Services

Similar to the Staffing Domain, cut points for the quality measures were based on information available in December 2011 and will be maintained for a two-year period.

Once the quality measure score is computed, the Five-Star rating is assigned for the Quality Measure Domain using the point thresholds provided in Table 10.

Table 10

*Five Star Cut Points for MDS Quality Measure Domain Score*

<table>
<thead>
<tr>
<th>1 Star</th>
<th>2 Stars lower</th>
<th>2 Stars upper</th>
<th>3 Stars lower</th>
<th>3 Stars upper</th>
<th>4 Stars lower</th>
<th>4 Stars upper</th>
<th>5 Stars</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤355</td>
<td>356</td>
<td>435</td>
<td>436</td>
<td>507</td>
<td>508</td>
<td>615</td>
<td>≥616</td>
</tr>
</tbody>
</table>

Source: Centers for Medicare & Medicaid Services
**Overall SNF rating (composite measure).** Based on the Five-Star rating for the Health Inspection Domain, Staffing Domain, and the Quality Measure Domain, an overall SNF rating is assigned to each SNF using the five-step process described in Table 11. The overall rating cannot be more than five stars or less than one star.

CMS has provided rationales for the star limitations and upgrades defined in the five-step process. The rationale for SNFs receiving an upgrade at Step 2 is that the criteria for the staffing rating are quite stringent. However, requiring that the staffing rating be greater than the health inspection rating in order for the score to be upgraded ensures that a SNF with four stars on health inspections and four stars on staffing does not receive a five-star overall rating.

The rationale for limiting upgrades in Step 4 is that two self-reported domains should not significantly outweigh the rating from actual onsite visits from trained surveyors who may have found serious quality issues. Also, the health inspection rating is heavily weighted toward the most recent annual inspection, meaning that a one-star health inspection rating reflects both a serious and recent finding.

The rationale for limiting the overall rating of a special focus facility in Step 5 is that the three domains are weighted toward recent results. This means that the domains do not fully take into account the history of some SNFs that fall in and out of compliance with federal safety and quality requirements. Once a SNF graduates from the SFF initiative by sustaining compliance for 12 months, the overall three-star cap on the overall rating is removed.

As seen in the five-step process, the overall SNF rating does not assign specific weights to the Health Inspection, Staffing, and Quality Measure Domains. The health
inspection rating is the most important dimension in determining the overall rating, but a
SNF’s performance on the Staffing and Quality Measure Domains can move a SNF’s
overall rating two stars higher or lower than their health inspection rating.

Table 11

_Five-Step Process for Assigning Overall Five-Star Rating_

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Start with the health inspection rating.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Add one star to the Step 1 result if the staffing rating is four or five</td>
</tr>
<tr>
<td></td>
<td>stars and greater than the health inspection rating. Subtract one star if</td>
</tr>
<tr>
<td></td>
<td>the staffing rating is one star. The overall rating cannot be more than</td>
</tr>
<tr>
<td></td>
<td>five stars or less than one star.</td>
</tr>
<tr>
<td>Step 3</td>
<td>Add one star to the Step 2 result if the quality measure rating is five</td>
</tr>
<tr>
<td></td>
<td>stars. Subtract one star if the quality measure rating is one star.</td>
</tr>
<tr>
<td>Step 4</td>
<td>If the health inspection rating is one star, the overall rating cannot be</td>
</tr>
<tr>
<td></td>
<td>upgraded by more than one star based on the staffing and quality measure</td>
</tr>
<tr>
<td></td>
<td>ratings.</td>
</tr>
<tr>
<td>Step 5</td>
<td>If the SNF is a special focus facility (SFF) that has not graduated, the</td>
</tr>
<tr>
<td></td>
<td>maximum overall rating is three stars.</td>
</tr>
</tbody>
</table>

_Source: Centers for Medicare & Medicaid Services_

**Issues with Current Quality Indicators.**

Measuring SNF quality and comparing providers’ performance has emerged as
the most hopeful strategy for holding them accountable for the care they provide (Jencks,
1994). Quality measurement, performance monitoring, and quality improvement are a
constant refrain in SNFs (Cohn, Corrigan, and Donaldosn, 2000). There are a large and
confusing number of quality indicators prevalent in SNFs. Individual quality indicators
(e.g., survey deficiencies) and sets of quality indicators (e.g., Five-Star ratings) are encumbered by several issues.

**Parsimony.** There is no single quality indicator that represents overall quality of a SNF. The deficiency rate resulting from annual and complaint survey investigations is the most frequently used and mentioned quality indicator, but it does not encompass overall quality (Stevenson, 2006). The Five-Star ratings provide a snapshot or simplified look at how SNFs compare on quality, but these ratings comprise several quality indicators (Castle et al., 2010).

The availability and use of multiple quality indicators have limitations. Mor (2005) found that, empirically, quality indicators have orthogonal relationships. That is, multiple dimensions of quality are thought to exist. Mor (2005) hypothesized that this occurs because SNFs provide care on multiple dimensions and may not be consistent in the quality of care provided for each dimension. For example, SNFs provide medical care and care for social needs. A SNF may provide quality medical care but be lacking in the social dimension of care.

Thus, the number of needed quality indicators is a vexing issue (Castle et al., 2010). A narrow focus on a single or small number of quality indicators may lead to incorrect conclusions about quality at a SNF. A focus on a moderate or large number of quality indicators introduces the risk of confusion and may also be misleading or incomplete.

Finally, a standard set of quality indicators assumes that all SNFs have the same goals and seek to treat the same types of patients. There are significant differences between SNFs in the types of patients they serve (Mor, et al., 2003). For example, some
SNFs that collaborate with hospice providers for terminal care management may not seek to achieve functional improvement. As a result, quality indicators assessing quality improvement would be inappropriate to use for some of the residents residing at SNFs that collaborate highly with hospice providers (Miller, Gozalo, and Mor, 2001).

**Measurement Issues.** For relatively rare outcomes, quality indicators have limited ability (power) to detect real differences in quality between SNFs. The standard errors for rare events are large, giving rise to several issues. First, the true quality level falls within the standard error, so the reliability of a single measure is questionable. Second, this is compounded when comparing one SNF to another. The large error bands make it difficult to differentiate whether one SNF has a truly better or worse quality level than another SNF.

Deficiency citations resulting from a complaint survey investigation can be classified as a rare outcome. Stevenson (2006) found that one-third of SNFs had zero complaint survey deficiencies and nearly half had only one deficiency. Also, over 80% of SNFs had less than five complaint survey deficiencies in a given time period. Despite the low number of citations, quality assessed by this indicator shows that SNFs with less than two citations differed significantly (p < .01) from SNFs with zero deficiency citations (Stevenson, 2010). Consumers could be confused by this outcome and be led to believe that a rare outcome (i.e., complaint survey deficiencies) can differentiate the quality provided between two SNFs with a similar number of deficiencies.

A further measurement issue is the assumed linearity of quality indicators. For example, a 10% rate of a quality indicator (e.g., pressure ulcers) may not represent twice the quality problem of a 5% rate. Also, linearity assumes full use of the scales such that
0% and 100% are possibilities. This is improbable for many quality indicators. For example, Lyder (2003) found that pressure ulcer rates less than 2% are considered improbable. This means that the implied scales are not clinically achievable. Castle and Enberg (2005) reported that the functional form of quality indicators is often unexpected and nonlinear.

A final measurement issue is ascertainment bias. Ascertainment bias is a type of detection bias where higher-quality SNFs may be actively looking for problems compared to lower-quality SNFs. Sangle (2005) hypothesized that lower-quality SNFs may have high staff turnover or high agency staff use, which could result in inadequate documentation. As such, higher-quality SNFs may have systematically higher-quality indicator rates, and lower-quality SNFs may have systematically lower indicator rates.

Ascertainment bias has been a measurement concern with both MDS data that is used to populate many of the Five-Star rating domains and deficiency citations resulting from annual and complaint survey investigations. Rahman and Applebaum (2009) report issues with interrater variability in MDS data, thus influencing the usefulness of these data. Also, deficiency citations differ between state survey teams as many states do not emphasize the same deficiency citations. States also differ in how aggressive they are in the use of deficiency citations in general (Stevenson, 2006).

**Consumer-determined quality.** The information used to calculate a SNF’s Five-Star ratings are often criticized because they focus almost entirely on medical issues. The medical focus does not include dimensions of quality that consumers also value (e.g., satisfaction). The “Making the Quality Connection” report (2005) by My InnerView, a leading provider of SNF patient and family satisfaction assessment, noted that the MDS-
based quality indicators used in the Five-Star ratings are an important advance in the measurement of SNF clinical quality, but there is a need to improve quality as defined by the experiences of patients and family members. Leading professional organizations (e.g., the American Health Care Association (AHCA)) recommend that a customer satisfaction survey tool be in place, as well as a process for conducting and analyzing satisfaction survey data as least annually as a baseline.

Use of patient and family satisfaction scores represents one means for including a consumer “voice” as a quality indicator (Sangle et al., 2007). Some states (e.g., Ohio) currently include satisfaction scores as a quality indicator (Ejaz, Straker, Fox, and Swami, 2003). Satisfaction information is often reported on state-run websites separate from the Nursing Home Compare website and may also include additional quality indicators not found in the Five-Star rating domains. The collection of satisfaction information is currently not a common practice due to the expense in collecting this data (Sangle, et al., 2007).

The introduction of Consumer Assessment of Healthcare Providers and Systems (CAHPS), a family of survey instruments designed to capture and report people’s experiences obtaining health care, could soon add satisfaction as a consistent dimension of quality that SNFs are required to assess and report (Sangle, et al., 2007). The CAHPS process has been introduced and federally mandated in several areas of health care, including hospital (HCAHPS) and home health care (HHCAHPS).

The SNF setting has not yet been mandated to implement CAHPS for nursing homes (NHCAHPS). The NHCAHPS surveys include two separate instruments for nursing home residents: one for those who live in a nursing home (Long-Stay Resident
Survey) and another for those who have been discharged after a short stay (Discharged Resident Survey). The NHCAHPS surveys also include an instrument for gathering information on the experiences of the family members of residents in SNFs. The Family Member Survey asks respondents to report on their own experiences (not the resident’s) with the SNF and their perceptions of the quality of care provided to a resident living in a SNF. The Family Member Survey was developed to complement the Long-Stay Resident Survey.

The development of these instruments was jointly supported by the Centers for Medicare & Medicaid Services (CMS) and the Agency for Healthcare Research and Quality (AHRQ). Like all CAHPS surveys, the instruments are in the public domain. In March 2011, the National Quality Forum (NQF) endorsed all three of these instruments as measures of nursing home quality, with the intention of mandating the use of NHCAHPS by 2014.

With the introduction of NHCAHPS, consumers will have another source of information to make comparisons between SNFs. When SNFs differ on satisfaction, it may impact a patient’s or family member’s decision on where to receive care. These comparisons justify the need for additional testing of the NHCAHPS surveys to include an assessment of the extent to which they are measures of patient and family satisfaction.

If the NHCAHPS Family Survey contains poorly constructed items, and family member ratings are summarized by an average of these items, there is no basis for knowing what is being measured and no basis for comparing the results between SNFs. To make more reliable and valid statistical inferences, it is necessary to construct a linear measure from the NHCAHPS Family Survey items before any statistical conclusions can
be drawn from family member evaluations. The results obtained from the NHCAHPS survey have several high stakes implications that solidify the need to engage in the process of measure construction using Rasch Analysis (citation) to build an inferential and stable ruler of family satisfaction.

The Rasch Model.

The Rasch model provides a framework within which test developers can assess the extent to which their items adhere to the properties of inferential measures. The Rasch model involves examination of only one human attribute (i.e., family satisfaction) at a time (unidimensionality) on a linear “more than/less than” line of inquiry (Bond and Fox, 2007). The raw data are transformed into abstract, equalinterval units by log transformations of raw data odds and probabilistic equations.

Instruments calibrated using Rasch modeling act as variable rulers. Variable rulers allow the placement of both items and persons on the same linear dimension. For example, if a family member is said to display a certain level of dissatisfaction with the SNF, the specific set of reports corresponding to that level of dissatisfaction is intuitively displayed in the same manner as measuring the length of an object using a ruler.

Additional Rasch statistics assist in the evaluation of the constructed ruler. Rasch statistics include fit statistics, separation statistics, item and person reliability estimates, and identification of gaps in the measure where items or persons are not well-targeted. Rasch fit statistics determine whether each item contributes to the measurement of the construct by assessing the extent to which an item or person performs as expected. An effective measure for assessing any given trait should consist of a mix of difficult items (those that are only endorsed by the most able/satisfied respondents), mid-range items
(those that are endorsed by all respondents except those with the lowest ability/satisfaction), and easy items (those that are endorsed by nearly everyone) (Fendrich, Smith, Pollack & Mackesy-Amiti, 2008).

Next, Rasch separation statistics estimate the ability of the items to assess different levels of the measure on a less-to-more continuum, and to identify the number of subgroups the set of items can discriminate. Furthermore, item and person reliability indices estimate the replicability of item placements and person ordering if the instrument were given to another appropriate sample or if the sample were given another set of items measuring this construct.

Finally, Rasch analysis shows gaps in the measure by identifying items and persons that are not well-targeted and hence yield larger than desirable error estimates. These gaps provide feedback on how well the instrument is actually measuring what it is supposed to measure and what might be done to further improve it (Bond & Fox, 2007).

There are several large-scale outcomes resulting from the measurement and reporting of all quality indicators used to assess SNF performance. Steps used to help ensure the reliability and validity of all quality indicators reported to the public improve the usefulness of this information when it is used in pay-for-performance programs and SNFs participate in Accountable Care Organization (ACO) agreements with health care systems.

**Outcomes of Quality Indicators.**

Performing well on quality indicators has several important outcomes to the continued growth and success of a SNF. First, SNFs are now being subjected to pay-for-performance initiatives that can have a direct impact on them financially through
increased or decreased reimbursement rates for services provided to patients. Second, with the introduction of the Affordable Care Act (ACA) in March 2010, hospitals are being provided incentives to create Accountable Care Organizations (ACOs). Within an ACO system, hospital executives are entering into contractual agreements with SNFs as part of a shared savings program to promote accountability of providers to patient populations and to coordinate services under Medicare. A hospital’s choice of SNF partners and, therefore, where it refers most patients is often dependent on the SNF’s performance on quality indicators.

**Pay for performance.** One strategy for improving the quality of care provided in SNFs is to link health care spending to quality and efficiency through pay-for-performance programs. The concept of pay for performance has become increasingly popular in an effort to improve quality and save money (Levenson, 2006). The underlying idea is to reimburse providers at least partially based on their performance in attaining specific goals or meeting specific requirements.

The pay-for-performance approach is in contrast to fee schedules of flat rates per service, where reimbursement is the highest when the most services are rendered, regardless of improved patient outcomes (Briesacher, et al., 2009). Hospitals, physicians, home health agencies, and laboratories started to see the pay-for-performance reimbursement approach become reality in 2004 when the Medicare Payment Advisory Commission (MedPAC) staff proposed that Congress should establish quality incentives, which could determine part of their Medicare payments for services rendered.

The outcome of the MedPAC proposition was the introduction of value-based purchasing (VBP) in October 2012. VBP is a payment reform under which hospitals and
other health care providers are given bonuses based upon their performance against quality measures (Donahue, 2012). The move is intended to help CMS move from being a passive bystander to an active buyer of what its officials have deemed higher quality care. VBP is not a collaborative approach among hospitals to improve care but a competition in which every hospital is pitted against the entire market. It is also a zero-sum financial game. Zero-sum means there will be winners and losers, with the entire cost-neutral program funded by extracting money from the worst performers to financially reward the best (Press and Fullman, 2011).

Eight of the quality measures used in the VBP program were taken from the HCAHPS patient surveys. HCAHPS scores account for 30% of a hospital’s final VBP score, while clinical measures are weighted at 70%. This means that patients’ perceptions of the care provided are an important part of the reimbursement a hospital receives through VBP.

The health care reform law also establishes the development of a similar VBP program for SNFs. CMS began sponsoring a pay-for-performance initiative through its Nursing Home VBP demonstration in the summer of 2009 (Brierecher, et al., 2009). Though the Nursing Home VBP program has yet to be implanted, some SNFs have been required to participate in pay-for-performance programs through their state-run Medicaid programs. From 2002 to 2007, six states (Iowa, Minnesota, Kansas, Georgia, Ohio, and Oklahoma) have used a pay-for-performance program in their SNFs (Arling, Job, and Cooke, 2009).

Levenson (2006) argues that if pay for performance is ever going to achieve a positive impact on quality of care, it must begin with valid and accurate measures that are
carefully selected to be related specifically to the desired behavior change. Quality measures must be specific to actionable issues, rather than a lengthy list of good ideas. This statement is especially important in the development of a patient satisfaction measure (e.g., NHCAHPS). SNF management teams will need the NHCAHPS feedback reports to identify stable and meaningful benchmarks in order to take action from the results. Without this information, SNF management teams may utilize their own benchmark (e.g., the company average score), under which being “above average” may still indicate poor performance.

**Accountable care organizations.** An accountable care organization (ACO) is a health care organization characterized by a payment and care delivery model that seeks to tie provider reimbursements to quality metrics and reductions in the total cost of care. An ACO is a network of doctors, hospitals, SNFs, and home health agencies that shares responsibility for providing care to patients. ACOs make providers jointly accountable for the health of their patients, giving them financial incentives to cooperate and save money by avoiding unnecessary tests and procedures.

The Affordable Care Act included incentives for the creation of ACOs. Congress established the Shared Savings Program in the Affordable Care Act to promote accountability of providers to patient populations and to coordinate services under Medicare, as well as to encourage providers to make investments in infrastructure and to design care processes for high-quality, efficient service delivery.

There are three goals stressed under the Shared Savings Program. First, provide better care to patients with respect to safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity. Second, provide better health for populations through
preventive service and education. Third, decrease the cost of health care and eliminate waste in the system. CMS is seeking to move the health care industry toward a patient-centered approach through increased involvement of patients in decisions about their care and requiring the collection and reporting of patient satisfaction data. ACOs will receive shared savings only if they can meet quality standards related to these three goals.

Hospitals, primary care providers, and other physicians are in charge of an ACO. Depending on the level of integration and size of an ACO, hospital providers will look to involve SNF organizations to help provide coordinated care to patients, align incentives, and lower overall health costs. As a result, SNFs that show the best performance in quality will often be asked to join the ACO. This will give the SNF increased access to referrals from the hospital systems with which they have ACO agreements, thus providing them a steady flow of Medicare patients. Patients who are receiving care in an ACO system are able to receive care at a SNF outside of the ACO system, but the referral will be made first to the SNF within the ACO.

One of the primary quality measures hospitals will examine when choosing a SNF ACO partner is patient satisfaction. In a 2012 HealthLeaders Media survey, 84% of hospital executives placed patient satisfaction in their top three priorities (Zeis, 2012). Hospital executives will also want to feel confident that the patients they refer to a SNF for additional care are receiving care at a SNF that also puts patient satisfaction as a high priority. This means that the survey used to assess satisfaction with a SNF will need to provide hospital executives with a meaningful benchmark when choosing SNF partners in an ACO.
Chapter Three
Methods and Procedures

Research Participants.

The population used in this study included family members of current skilled nursing facility (SNF) residents who completed the NHCAHPS Family Survey. The data were collected from SNFs operated by a large provider organization in the United States. The data collection process was part of the 2012 regularly scheduled family satisfaction survey the organization conducts every other year. Surveys were mailed to 16,649 family members and 5,189 surveys were returned for a 31% response rate. This response rate was close to past response rates for family satisfaction surveys conducted by the organization.

The obtained 31% response rate was a reasonable response rate for this project. For a 95% confidence level (which means that there is only a 5% chance the sample results differed from the true population average), a good estimate of the margin of error (or confidence interval) is given by $1/\sqrt{N}$, where $N$ is the number of participants or sample size (Niles, 2006). This means that the margin of error with 5,189 returned surveys was .01. A 31% response rate at an individual facility level would have been problematic for analyses because the margin of error would have been well above .05.

Instrumentation.

The NHCAHPS Family Survey was used to assess family satisfaction (see Appendix A for a copy of the Family Survey). The NHCAHPS Family Survey includes questions for respondents to report about their own experiences (not the resident’s) with
the SNF and their perceptions of the quality of care provided to a resident living in a SNF.

The NHCAHPS Family Member Survey contains 50 items. Not all 50 items are used for public reporting of NHCAHPS scores. The CAHPS team selected a final set of 21 substantive items in 4 composite domains (see Table 2 for a definition of each composite domain): Meeting Basic Needs; Nurses/Aides Kindness & Respect towards Resident; Nursing Home Provides Information/Encourages Respondent Involvement; and Nursing Home Staffing, Care of Belongings & Cleanliness. The only published research using the NHCAHPS Family Survey instrument comes from the American Institute of Research (AIR) Final Report (2008) reported Cronbach’s alpha internal consistency reliability estimates that ranged from 0.73 to 0.81 for the composite domains.

In addition to the 21 items used in the composite domains, the Overall Quality of Care item is also included resulting in 22 of the 50 survey items to assess family satisfaction. In the current study, a “Recommendation to Others” item was also included in the assessment of the NHCAHPS survey. This item was included due to the importance many SNF management team place on this item for business growth and development.

There are five unique rating scales used by the 23 items. The direction of each rating scale suggested that higher scores represented higher levels of family satisfaction. The first rating scale (Never, Sometimes, Usually, Always) was used by 13 items. The second rating scale (Never, Once or Twice, and Three or More Times) was used by two survey items. Responses received for the two items that used this scale were reverse coded since a Never response was the desired response to these items. The third rating
scale (Yes or No) was used by six items. The fourth rating scale used a 0 to 10 scale (Worst Possible Care to Best Possible Care). This rating scale was used by one item (Overall rating of Care). The final rating scale (Definitely No, Probably No, Probably Yes and Definitely Yes) was used by the Recommendation to Others item.

The remaining 17 survey items include items about the resident (e.g., cognitive status), survey respondent demographic items and “Skip” items (see item 10 in Appendix A as an example of a “Skip” item). The eight “Skip” items on the survey were designed to determine whether a family member had experienced a specific type of event or incident while visiting the resident. The “Skip” items used a Yes or No rating scale. If the respondent answered Yes to a “Skip” item then he moved on to the next item on the survey. If a respondent answered No to a “Skip” item then he skipped the next survey item.

Data Management and Collection.

The data management process was managed by National Research Corporation (NRC). As an approved vendor of CAHPS surveys, NRC was required to adhere to the CAHPS Quality Assurance Guidelines. These requirements were established by the Centers for Medicare and Medicaid Services (CMS) to ensure CAHPS vendors have the necessary system resources, personnel training, and data security to ensure patient confidentiality.

The collection of data occurred over a two month period in the summer of 2012. The data collection process involved a series of phases that occurred according to a timeline set up for the study.
Phase 1: Facility notification of the survey. The first phase of the data collection process involved informing SNF management teams that a family survey was going to be conducted at the facility. The notification involved two forms of communication; email and webinar. First, an email was sent with information about the survey data collection dates, the process used to collect the data, a copy of the NHCAHPS Family Member Survey and steps they need to take to prepare for the survey. In addition to the email notification, SNF management teams were encouraged to attend a webinar that provided additional information about the survey and data collection process.

Phase 2: Assignment of survey recipient. SNF management teams were asked to assign a Customer Service Survey Recipient (CSSR) in the Point Click Care (PCC) patient information system. The CSSR was a custom contact type created in PCC for SNF management teams to identify the appropriate family member to receive a satisfaction survey. Additional contact types in PCC included power of attorney, emergency contact, Accounts Receivable (A/R) representative, etc. Typically, the family member chosen was also listed as one of these additional contact types.

If a CSSR was not assigned for a resident then the A/R Representative contact type was assigned as the survey recipient for that resident. The A/R Representative contact type was used as the default because this contact type typically has a completed address and phone number listed since he/she is responsible for bills sent by the facility.

Phase 3: Survey notification and distribution. A survey notification postcard was sent to each person on address list. The use of the postcard was designed to help increase response rate to the mail survey. One week following the distribution of the pre-
notification postcard, the data collection process occurred. Data collection followed CAHPS protocol that allowed 6 weeks for the survey to be returned.

Theoretical Framework for Measure Construction.

The purpose of this study was to use the NHCAHPS Family member survey as a beginning point in the development of a measure of family member satisfaction.

Measurement begins with the idea of a variable (construct) or a ruler along which persons can be positioned and the intention to mark off this line in equal units so that distances between points on the ruler can be compared (Wright and Masters, 1982). A line of inquiry or a line of observations, such as an increasing degree of family satisfaction that provides an operational definition of family satisfaction, does not immediately provide an effective ruler.

Linearity requires that the obtained numbers through observations preserve not only the order of the objects and instruments, but also the order of their differences (Wright and Masters, 1982). To derive numbers for items and persons that enable quantitative comparisons that can be maintained over a useful range of generality the following basic requirements must be met (Wright and Masters, 1982):

a. The reduction of experience to a one dimensional abstraction, such as family satisfaction in this study;

b. More or less comparisons among persons and items;

c. The idea of linear magnitude inherent in positioning items and persons along a line; and
d. A unit determined by a process which can be repeated without modification in the different parts of the measurement continuum – a theory or model for how persons and items must interact to produce useful observations.

If the basic requirements are satisfied, the construct of the measurement contains an abstract of a single line of one-dimension inquiry, along which items and persons can be located on the basis of observation which add up. The same values are expected for a particular measure no matter which items are used as long as they define the same construct.

**Rasch rating scale model.** The Rasch model is a psychometric model that helps create measures. The Rasch model is built upon the requirement that the person and item parameters can be estimated independently of each other (Wright and Stone, 1979). The Rasch model follows two basic specifications: “*Persons who are more able or more developed have a greater likelihood of correctly answering all the items in the observation schedule*”; and “*Easier items are more likely to be answered or reached correctly by all persons*” (Wright & Stone, 1979; Bond & Fox, 2007). Based on these two specifications, the Rasch model provides the necessary objectivity for the construction of a scale that is separable from the distribution of the attribute of the persons it measures.

In operation, the Rasch model transforms raw data into abstract, equal-interval scales. Linearity or equality of intervals is achieved through the log transformation of raw data into odds ratio:

\[
B_p (\text{person ability}) = \ln \frac{P}{1-P}
\]
\[ D_i (\text{Item difficulty}) = \ln \frac{P}{1 - P} \]

where \( p \) is the proportion correct or proportion passed. To make the item scores and the person scores run in the same direction, the direction of the item scores and the person scores should be reversed in the transformation. Thus, the observed item score \( S \) can be transformed into “logit” score \( D_i \) through the following logit transformation:

\[ D_i = \ln \frac{N - i}{S} \]

where \( N \) is the number of participants in the sample (Wright & Masters, 1982). This transformation changes high item raw scores \( S \) into low logit scores \( D_i \), so that on the scale the items that are easiest to endorse have lowest logit scores, indicating the least difficulty.

The rating scale model was developed by Andrich (1978). Andrich added the concept of thresholds or steps to the basic Rasch model. For an attitude item with the options ranging from Never to Always, “completing the k’th step” is choosing the k’th option over the (k-1)’th option in response to the item.

A further step can be taken only if the first step has been completed. Thus, choosing “Usually” means having chosen “Sometimes” over “Never” (first step taken). Also, the respondent must choose “Usually” over “Sometimes”, but failed to choose “Always” over “Sometimes”. According to Wright and Masters (1982), the relative difficulties of the “steps” in a rating scale item are intended to be governed by the fixed set of rating points, and as the same set of rating points is used with every item, it is assumed that the relative difficulties of the steps in each item should not vary from item to item. Thus, the only difference between items is the difference in their location along
the line of inquiry, and the person’s observed score is the interaction of person ability, item difficulty, and step difficulty.

The danger of an unfair decision, by a family member, to place a resident in one SNF over another that might follow from comparing uncorrected raw scores is minimized through this model. When applied to SNFs, this model places facilities in an objective frame of reference, thus enabling adjustments to the facility assessment based on inconsistencies in family member ratings.

Rasch analysis separates the estimates of family member ratings and item difficulty so that each set of estimates is independent of the distributions of the other. This approach will reduce the probability that a SNF that provides high quality care will receive low ratings due to unlucky encounters with severe assessors (i.e., family members) or that SNFs that provide low quality care will receive high ratings due to lucky encounters with assessors.

Data Analysis.

The data were analyzed with the Rasch rating scale model (Andrich, 1978) in the WINSTEPS (Linacre and Wright, 2004) software. The survey was examined across aspects of error, targeting, separation, hierarchical order, fit statistics, and dimensionality. These aspects provided important information on the reliability and validity of the NHCAHPS Family Member Survey.

Reliability. Reliability refers to the reproducibility of results and how likely the measures are due to chance (Linacre, 2006). Reliability is calculated as the ratio of the true variance and the observed variance. The Rasch model analyzes reliability in two indices: person and item. The person reliability index was examined as an indicator of
the expected replicability of person ordering if the sample of persons were given another set of items measuring the same construct. The item reliability index was examined as an indicator of the expected replicability of item placement along the variable ruler if the same items were given to another sample with comparable ability levels. Item and person reliabilities were examined using the entire respondent pool and through random samples of 500 respondents.

Reliability was also assessed through an examination of person and item separation. Separation is a measure of the spread of the estimates relative to their precision (Linacre, 2006) and is calculated as the ratio of “true” (sample) standard deviation to the error standard deviation. Person separation was used to describe how well the survey identified individual differences. Item separation was used to determine how well the survey separates the items. The higher the separation, the more confidence can be placed on the replicability of item placement across samples (Bond and Fox, 2007).

Separation can be thought of as the number of levels into which the sample of items and persons can be separated. For a survey to be useful, separation should exceed 1.0, with higher values of separation representing greater spread of items and persons along a continuum. Lower values of separation indicate redundancy in the items and less variability of persons on the construct (i.e., family satisfaction).

Rasch person and item separation statistics (G) were examined for the entire survey and for each composite measure to determine the level of distinction possible among persons and items along the family satisfaction construct. Rasch separation statistics were then transformed into a strata index, which determines the number of
statistically different levels of person ability that are distinguished by the items. Strata were calculated using the formula \( [4G+1]/3 \) (Fisher, 1992). A separation of 2.0 (i.e., identifying three strata) was used as the minimum acceptable value (Wright and Masters, 1982).

Finally, the Rasch model specifications require that a properly functioning survey should contain a reasonable spread of items from less to more difficult and that those items match the relative level of endorsement by survey respondents. Specifically, items should be neither too hard nor too easy in their elicitation of responses. This is a concept referred to as targeting. If the items are all too easy to affirm (i.e., most respondents select “Always”) or too easy to deny (i.e., most respondents select “Never”) then these items fail to assist in the development of reliable outcomes because they fail to differentiate from those who agree from those who do not. The distribution of items and persons on the person-item maps were examined to assess the targeting of the survey. If too many respondents outscore the items then it could indicate a need to add additional (more difficult) items to estimate the abilities of these respondents more precisely.

**Validity.** Construct validity was assessed through an examination of rating scale structure, item fit statistics, and the dimensionality of the survey instrument. Evaluation of the functioning of the rating scale categories was completed through an examination of common Rasch scale diagnostics. These diagnostics provided information about how respondents used each of the five different rating scale types used in the survey. The first diagnostic used to assess rating scale functioning was the step thresholds. Step thresholds are the estimated difficulties in choosing one response option over another (e.g., choosing *always* over *sometimes*). Thresholds were examined to determine whether they increase
monotonically and to determine the distance between thresholds. The magnitude of the step thresholds used the recommended guideline of at least a 1.4 logit increase to show distinction between categories (Linacre, 1999).

Another rating scale diagnostic that was used entailed a visual inspection of the probability curve graph to ensure useful distinctions among response categories. The probability graph curves show the probability of choosing a given rating scale category for every place along the measured construct. Useful categories were determined as those with high probabilities that spanned a distinct portion of the construct (Bond and Fox, 2007). Categories that overlap too much with adjacent categories are typically not helpful in defining a distinct point along the construct. The probability curves visually display the same information as the step thresholds.

Any of the five rating scale categories that had rating scale diagnostics that did not meet the criterion noted above were further examined. Inconsistent use of a rating scale impacts item fit and placement. Inconsistent use of a rating scale was resolved by collapsing categories and reanalyzing the data to reassess the optimal number of categories for the data. Rating scale diagnostics were conducted first with the entire data set and then reassessed with five random samples of 500 respondents.

Next, fit statistics were examined as an indicator of how well each item fit within the underlying construct. Items can “overfit” and “underfit”. To determine whether any of the items on the NHCAHPS Family Member survey captured something qualitatively different from family satisfaction, in-fit mean squares were examined, with the value of 1.4 used as the cutoff for rating scales (Bond and Fox, 2007). Item redundancy was investigated by the out-fit statistics, on the basis of the same criterion (1.4).
In addition to examining fit statistics, point-measure correlations were also reviewed. Point-measure correlations provided information on how each item correlated with the latent variable (family satisfaction). Items that misfit through an examination of the fit statistics and those with a small or even negative point-measure correlation were examined to determine if there was the presence of another underlying construct being measured by the survey. Second, a Rasch principle contrasts analysis (RPCA) was conducted. The RPCA was used to uncover the presence of multiple dimensions. The benchmark of at least 60% of the variance being explained was used as the criterion for unidimensionality in the RPCA.

Additional evidence of construct validity of the NHCAHPS Family Member survey was obtained through a review of the survey items ranked by difficulty as displayed on the general keyform map. The general keyform map illustrates the ordering of the satisfaction items from least (top) to most (bottom) agreeable, on an equal-interval scale. An informal qualitative analysis was conducted on the groupings of items on the general keyform map. Examination of the item ordering and groupings provided insight into meaning of obtained ordering of item.
Chapter Four

Results

Diagnostic Analyses.

Analysis of the NHCAHPS Family Member survey began with a series of Rasch diagnostic tests designed to assess reliability and validity of the constructed measure. The first set of diagnostic analyses used the entire data set (n=4569). Smaller random subsamples (n=500) were used also used to replicate findings across cross-sections of data.

Reliability analyses. Rasch person and item separation statistics were 2.75 and 21.62, respectively. The person separation of 2.75 translated into four statistically distinct strata. This means that persons were measured on a continuum rather than a dichotomy. Item separation was 21.62 which translated into 29 distinct strata. The larger item separation was expected since there are there were only 23 items and 4,569 people. Separation is affected by sample size. This means that the item separation calculated from the smaller random subsamples may provide a more appropriate estimate of the item separation.

Person and item reliabilities were also high, corresponding to alpha levels of .88 and 1.0. This means that the replicability of person ordering for a parallel set of items given to this sample and item placement along the pathway if these items were given to a similar group of people was high. Reliability estimates are also dependent on sample size meaning the reliability estimates calculated from the smaller random samples will provide a more appropriate estimate of reliability.
Scale category analyses. The NHCAHPS Family survey used five separate rating scales. The scale category analyses examined the functioning of each rating scale separately.

Rating scale type 1. The first rating scale type included items that used a 4-point (never, sometimes, usually and always) scale. Table 1 shows the diagnostics for the 13 items that used this rating scale.

Table 12

Rating Scale Type 1 Diagnostics

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>1201</td>
<td>-.86</td>
<td>1.26</td>
<td>1.57</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Sometimes</td>
<td>5539</td>
<td>.21</td>
<td>.95</td>
<td>.99</td>
<td>-1.94</td>
<td>.03</td>
</tr>
<tr>
<td>Usually</td>
<td>19579</td>
<td>1.62</td>
<td>.89</td>
<td>.78</td>
<td>-.30</td>
<td>.02</td>
</tr>
<tr>
<td>Always</td>
<td>24976</td>
<td>3.37</td>
<td>.96</td>
<td>.97</td>
<td>2.24</td>
<td>.01</td>
</tr>
</tbody>
</table>

As Table 12 indicates, the average measure for the rating scale categories advanced monotonically from never to always. The step thresholds also met the recommended guideline of at least a 1.4 increase between rating scale categories meaning that respondents did reliably distinguish between the rating scale categories. In addition, the in-fit and out-fit mean squares met the criterion of being less than 2.0 (Linacre, 1999).

The probability curves for each rating scale category are displayed in Figure 2. Examination of the probability curves revealed that each response category had a distinct peak in the probability curve graph. This means that each rating scale category was the
most probable response for some portion of the measured variable and each rating point is distinctly different from the other rating points. Figure 2 also shows that the average measure for each rating point increases monotonically.

Figure 2. Rating Scale Type 1 Probability Curves

The graph in Figure 2 also illustrates the probability of a responding to any particular category, given the difference in estimates between any person ability and item difficulty. For example, if a person’s ability were 1 logit lower than the difficulty of the item (-1 on the x-axis), that person’s probability of choosing a four (always) was close to zero. The probability of that person choosing a one (never) or a three (usually) was close to .2. The probability of that person choosing a two (sometimes) was .55. This means
that persons with ability estimates one logit below the item difficulty had a most probable response of two (sometimes) to the item. For respondents with ability estimates higher than the given item difficulty (e.g., +1 on the x-axis), the most probable response was three (.65 probability).

**Rating scale type 2.** The second rating scale type included items that used a 3-point (never, once or twice, and three or more times) scale. Responses to the two items that used this scale were reverse coded since a never response is the desired response to these items. Table 13 shows the diagnostics for this rating scale.

Table 13

**Rating Scale Type 2 Diagnostics**

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three or More Times</td>
<td>1355</td>
<td>-.23</td>
<td>1.76</td>
<td>2.62</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Once or Twice</td>
<td>2213</td>
<td>.70</td>
<td>1.36</td>
<td>2.13</td>
<td>-.67</td>
<td>.04</td>
</tr>
<tr>
<td>Never</td>
<td>3771</td>
<td>1.63</td>
<td>1.78</td>
<td>1.98</td>
<td>.67</td>
<td>.03</td>
</tr>
</tbody>
</table>

The average measure and the step thresholds for the rating scale categories advanced monotonically from three or more times to never. The step threshold increase between rating scale categories two and three was 1.34, which is slightly below the recommended guideline of a 1.4 increase. Also, the out-fit mean square statistics were above the 2.0 threshold for two of the response categories and nearly above 2.0 for all
categories. This is an indication that these rating scale points were introducing noise into the measurement process.

The probability curves for each rating scale category are displayed in Figure 3. Examination of the probability curves revealed that each response category had a distinct peak in the probability curve graph. Rating scale category two (\textit{once or twice}) did have a distinct peak but covered only a small portion of the variable (from \(-.4\) to \(+.6\) logits) meaning that response categories one and three were the most probable responses for the largest portions of the variable.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{rating_scale_type2_prob_curves.png}
\caption{Rating Scale Type 2 Probability Curves}
\end{figure}
**Rating scale type 3.** The third rating scale type would not technically be considered a rating scale since the response options were a dichotomy (yes or no). For ease of discussion about the five response options used on the NHCAHPS Family Survey, it will be listed as a rating scale type. There were six items on the survey that used this rating scale type. The average measure for rating scale categories increased monotonically. A *yes* response had an average measure of .20 logits and a *no* response had an average measure of 3.03 logits. The in-fit and out-fit mean squares were all below 1.0. These diagnostics indicated that respondents used this rating scale correctly.

**Rating scale type 4.** The fourth rating scale type included the Recommendation to Others item. This item used a 4-point (*definitely no, probably no, probably yes and definitively yes*) scale. Table 14 shows the diagnostics for this item.

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely No</td>
<td>171</td>
<td>-.64</td>
<td>.84</td>
<td>.82</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Probably No</td>
<td>358</td>
<td>.13</td>
<td>.65</td>
<td>.57</td>
<td>-1.07</td>
<td>.09</td>
</tr>
<tr>
<td>Probably Yes</td>
<td>1979</td>
<td>1.58</td>
<td>.72</td>
<td>.69</td>
<td>-.97</td>
<td>.06</td>
</tr>
<tr>
<td>Definitely Yes</td>
<td>1993</td>
<td>3.27</td>
<td>.74</td>
<td>.82</td>
<td>2.04</td>
<td>.04</td>
</tr>
</tbody>
</table>
The response categories followed the expected progression of rated levels, that is, they advanced monotonically from *definitely no* to *definitely yes*, as Table 14 indicates. However, as the step thresholds estimates in this table show, the two adjacent categories two (*probably no*) and three (*probably yes*) were not statistically different, .1 logits apart, indicating that respondents did not reliably distinguish between these categories.

![Table 14: Category Probabilities](image)

**Figure 4.** Rating Scale Type 4 Probability Curves

The probability curves for each rating scale categories are displayed in Figure 4. Examination of the probability curves revealed that highest probability for endorsement of category two (*probably no*) was approximately .35. Information found in Table 14 and from a visual inspection of the probability curves indicated that respondents did not reliably distinguish between response categories two and three. Respondents were unable
to make a distinction between response categories two and three and hence used them in an inconsistent manner. To improve the functioning of this rating scale, a collapsing strategy was employed (discussed below).

**Rating scale type 5.** The final rating scale type was used for only one item on the survey, the Overall rating of Care. This rating scale type had respondents rate the overall care at the SNF using a 0 (*worst possible care*) to 10 (*best possible care*) scale. There were no descriptions of overall care provided for the rating scale points one through nine. In the present study, none of the respondents used the zero rating scale point resulting in the scale for this item being 10-points rather than the intended 11-points. Table 15 shows the diagnostics for this rating scale.

As Table 15 indicates, the average measure for the rating scale categories advanced monotonically from *worst possible care* to *best possible care*. The step thresholds did not increase monotonically and did not meet the recommended guideline of at least a 1.4 increase between rating scale categories. This means that rating scale categories for this item were disordered. The in-fit and out-fit mean square statistics met the threshold of being less than 2.0.

Figure 5 provides the probability curves for each rating scale category. Examination of the probability curves revealed that each response category did not have a distinct peak in the probability curve graph. Also, rating scale categories two through seven had less than a .5 probability. Information found in Table 15 and Figure 5 indicated that respondents did not reliably distinguish between many of the response categories. To improve the functioning of this rating scale, several rating scale collapsing strategies were employed (discussed below).
Table 15

*Rating Scale Type 5 Diagnostics*

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>28</td>
<td>-1.21</td>
<td>1.30</td>
<td>1.50</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>25</td>
<td>-1.08</td>
<td>.97</td>
<td>1.00</td>
<td>-1.29</td>
<td>.22</td>
</tr>
<tr>
<td>Three</td>
<td>56</td>
<td>-.76</td>
<td>.76</td>
<td>.72</td>
<td>-1.87</td>
<td>.17</td>
</tr>
<tr>
<td>Four</td>
<td>85</td>
<td>-.43</td>
<td>.84</td>
<td>.83</td>
<td>-1.14</td>
<td>.13</td>
</tr>
<tr>
<td>Five</td>
<td>246</td>
<td>.06</td>
<td>.81</td>
<td>.77</td>
<td>-1.40</td>
<td>.10</td>
</tr>
<tr>
<td>Six</td>
<td>245</td>
<td>.52</td>
<td>.77</td>
<td>.76</td>
<td>.11</td>
<td>.07</td>
</tr>
<tr>
<td>Seven</td>
<td>564</td>
<td>1.04</td>
<td>.71</td>
<td>.63</td>
<td>-.19</td>
<td>.06</td>
</tr>
<tr>
<td>Eight</td>
<td>1177</td>
<td>1.86</td>
<td>.70</td>
<td>.68</td>
<td>.54</td>
<td>.05</td>
</tr>
<tr>
<td>Nine</td>
<td>1082</td>
<td>2.79</td>
<td>.72</td>
<td>.79</td>
<td>2.14</td>
<td>.04</td>
</tr>
<tr>
<td>Ten</td>
<td>978</td>
<td>3.89</td>
<td>.81</td>
<td>.86</td>
<td>3.11</td>
<td>.05</td>
</tr>
</tbody>
</table>

**Improving the Rating Scales**

Based on the diagnostic analyses, respondents used three of the five rating scale types as expected. Rating scale type four (Recommendation to Others) and rating scale type five (Overall rating of Care) both needed further refinement to be useful items in the NHCAHPS Family Member survey. Refinement of the rating scales involved collapsing response categories.
The first and foremost guideline in collapsing rating scale categories is that the points that are collapsed together must make sense (Wright, 1987). However, given that there were no rating scale labels for categories one through nine on the Overall rating of Care item, this wasn’t possible. This guideline was followed for the Recommendation to Others item scale as the category labels allowed for the collapsing decision to make sense. A second guideline for collapsing rating scale categories indicates that there should be an attempt to create a uniform distribution of responses (Linacre, 1995).

**Collapsing recommendation to others scale.** Currently, CMS does not report scores for the Recommendation to Others item independently or in the composite domain scores but it is an item many organizations use as a metric for customer loyalty.
(Reichold, 2003). Due to the importance organizations place on this item, a collapsing option was tested to improve the functioning of this rating scale type.

The refinement of this rating scale involved collapsing response categories one (definitely no) and two (probably no). These two response categories had the lowest frequency of responses, 171 and 358 respectively thus collapsing them helped to create a more uniform distribution of response. Also, collapsing categories one and two made sense because they both represented a respondent who indicated they were not likely to recommend the SNF to others. Collapsing these categories together changed the rating scale into a 3-point where rating scale category one was defined as “No” and the remaining two categories remained unchanged (probably yes and definitely yes). Table 16 shows the rating scale diagnostics for the refined 3-point rating scale.

Table 16

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>529</td>
<td>-.17</td>
<td>.71</td>
<td>.70</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>1979</td>
<td>1.53</td>
<td>.76</td>
<td>.68</td>
<td>-1.37</td>
<td>.05</td>
</tr>
<tr>
<td>Three</td>
<td>1993</td>
<td>3.23</td>
<td>.76</td>
<td>.79</td>
<td>1.37</td>
<td>.04</td>
</tr>
</tbody>
</table>

The average measure and the step thresholds for the rating scale categories advanced monotonically from rating scale categories one to three. The step threshold increases met the recommended guideline of being above 1.4 logits meaning that respondents reliably distinguished between the rating scale categories. In the original 4-
point rating scale structure, there was only a .1 logit increase from category two
(probably no) and category three (probably yes). The out-fit and in-fit mean squares
were all below 1.0.

Figure 6. Recommendation to Others 3-Point Rating Scale Probability Curves

Figure 6 provides the probability curves for refined rating scale. Examination of
the probability curves revealed that each response category did have a distinct peak in the
probability curve graph. Also, each of the rating scale categories had above a .5
probability of response and appeared to be the most probable response across an equal
section of the variable. Information found in Table 16 and Figure 6 indicates that
respondents were better able to distinguish between response categories in the refined 3-
point scale than with the original scale.
**Overall rating of care.** The Overall rating of Care item is considered one of the most important items on the survey as it is intended to provide a numerical assessment of the perceptions of overall care provided at the SNF (Sangle, et. al, 2007). The CAHPS survey designers developed this item with 11 rating scale categories to force variability in responses. This variability provides consumers with an indicator based on the raw score average of responses to show differences between SNFs on overall care.

Due to the importance of this item in the overall assessment of care provided at SNFs, four collapsing options were tested, each of which was chosen to test the maximum number of meaningful rating scale categories used by respondents. Each of these collapsing options was chosen through an examination of the probability curves displayed in graph in Figure 5 and the diagnostic information presented in Table 15. Appendix D contains a detailed description of each collapsing option and the resulting rating scale diagnostics. Table 17 provides a brief description of the collapsing options.

In addition to the rating scale diagnostic information for the four collapsing options presented in Appendix D, separation, reliability, variance explained by the measure and fit indices were compared between each collapsing option. Table 18 summarizes the four collapsing options used on the overall quality of care rating scale.
### Table 17

*Description of the Overall Rating of Care Collapsing Options*

<table>
<thead>
<tr>
<th>Collapsing Options</th>
<th>Description</th>
</tr>
</thead>
</table>
| Option A (6-point) | Categories 1, 2 and 3 collapsed together  
Categories 4 and 5 collapsed together  
Categories 6 and 7 collapsed together  
Categories 8, 9 and 10 unchanged |
| Option B (6-point) | Category 1 unchanged  
Categories 2, 3, and 4 collapsed together  
Categories 5 and 6 collapsed together  
Categories 7 and 8 collapsed together  
Categories 9 and 10 unchanged |
| Option C (5-point) | Categories 1, 2, 3, and 4 collapsed together  
Categories 5 and 6 collapsed together  
Categories 7 and 8 collapsed together  
Categories 9 and 10 unchanged |
| Option D (4-point) | Categories 1, 2, 3, and 4 collapsed together  
Categories 5 and 6 collapsed together  
Categories 7 and 8 collapsed together  
Categories 9 and 10 collapsed together |

The separation, reliability, and fit statistics remained unchanged regardless of the collapsing option employed meaning that changing the rating scale for this item did not have much impact on the precision of the entire measure. This means that this item may be capturing information already collected in the other survey items since this item was asking respondents to provide an overall rating. Collapsing Option A had the largest increase in variance explained (nearly 1%) when compared to the original 11-point rating scale meaning that this collapsing option did increase improve the unidimensionality of the survey.
Table 18

Summary of Changes in Person and Item Separation, Reliability, Variance Explained and Fit Indices for Overall Care Rating Scale Collapsing Options

<table>
<thead>
<tr>
<th>Rating Scale</th>
<th>Separation Person</th>
<th>Item</th>
<th>Reliability Person</th>
<th>Item</th>
<th>Mean Square In-Fit</th>
<th>Out-Fit</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original 11-point</td>
<td>2.75</td>
<td>21.62</td>
<td>.88</td>
<td>1.0</td>
<td>1.02</td>
<td>1.04</td>
<td>71.5%</td>
</tr>
<tr>
<td>Option A</td>
<td>2.75</td>
<td>22.60</td>
<td>.88</td>
<td>1.0</td>
<td>1.02</td>
<td>1.04</td>
<td>72.3%</td>
</tr>
<tr>
<td>Option B</td>
<td>2.74</td>
<td>22.22</td>
<td>.88</td>
<td>1.0</td>
<td>1.02</td>
<td>1.04</td>
<td>71.5%</td>
</tr>
<tr>
<td>Option C</td>
<td>2.74</td>
<td>22.64</td>
<td>.88</td>
<td>1.0</td>
<td>1.02</td>
<td>1.04</td>
<td>72.0%</td>
</tr>
<tr>
<td>Option D</td>
<td>2.62</td>
<td>22.19</td>
<td>.87</td>
<td>1.0</td>
<td>1.02</td>
<td>1.03</td>
<td>70.2%</td>
</tr>
</tbody>
</table>

Based on the results from the rating scale refinement for the “Overall rating of Care” and “Recommendation to Others” items, further analyses of this survey used the 6-point and 3-point refined rating scale types respectively. To further test the diagnostics of the survey, smaller random samples were taken to replicate the findings across cross sections of the data.

**Diagnostic Analysis with Random Samples**

Five random samples of 500 responses were taken from the 4,569 survey responses. Table 19 provides the item separation, person separation, reliability and variance explained for the five random samples and the entire data set.

The person separation for the five samples ranged from 2.67 to 2.75 and was consistent with the entire sample. This translates into 4 statistically distinct strata for respondents. The item separation ranged from 6.59 to 7.40. As expected, the item
separation was significantly lower than entire data set (22.60) because the separation is affected by sample size and the entire data set included 4,569 people while each random sample is only 500. The item separation for the random samples translates into approximately 9.41 distinct strata. Person and item reliabilities remained high for the random samples. The consistently high person and item reliabilities were an indication of the quality and usefulness of the items. The variance explained was above 71% for each of the random samples with the largest variance explained found in Sample three (75.2%).

Table 19

*Separation, Reliability and Variance Explained for Random Samples*

<table>
<thead>
<tr>
<th></th>
<th>Person Separation</th>
<th>Item Separation</th>
<th>Person Reliability</th>
<th>Item Reliability</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>2.70</td>
<td>6.81</td>
<td>.88</td>
<td>.98</td>
<td>71.1%</td>
</tr>
<tr>
<td>Sample 2</td>
<td>2.72</td>
<td>6.81</td>
<td>.88</td>
<td>.98</td>
<td>71.6%</td>
</tr>
<tr>
<td>Sample 3</td>
<td>2.75</td>
<td>7.40</td>
<td>.88</td>
<td>.98</td>
<td>75.2%</td>
</tr>
<tr>
<td>Sample 4</td>
<td>2.67</td>
<td>6.59</td>
<td>.88</td>
<td>.98</td>
<td>72.2%</td>
</tr>
<tr>
<td>Sample 5</td>
<td>2.71</td>
<td>6.95</td>
<td>.88</td>
<td>.98</td>
<td>71.8%</td>
</tr>
<tr>
<td>Entire Data Set</td>
<td>2.75</td>
<td>21.62</td>
<td>.88</td>
<td>1.0</td>
<td>72.3%</td>
</tr>
</tbody>
</table>

*Scale category analyses.* Rating scale types one, two and three continued to function across the five random samples in the same way they functioned when the entire data set was used. The average measure for these rating scale categories advanced monotonically, step thresholds met the recommended guideline of at least a 1.4 increase
between rating scale categories, and the in-fit and out-fit mean square statistics met the threshold of being less than 2.0.

Rating scale type four (Recommendation to Others), had been collapsed from a 4-point scale to a 3-point scale. The average measures and step thresholds increased monotonically across all samples. The step threshold increases were above 1.4 and the in-fit and out-fit mean squares were below 2.0 across all samples. This means that the revised rating scale continued to function in the same manner for a smaller sample as it did with the entire data set.

Rating scale type five (Overall rating of care), had been collapsed from a 0 to 10 scale to a one to six scale. The average measures and step thresholds increased monotonically across all samples. As seen with the collapsed scale in the entire data set, the step threshold increases continued to be around 1.0 logits with the largest increase (approximately 1.3 logits) found between rating scale points four and five. The in-fit and out-fit mean squares were also below 2.0 in all samples.

**Item Fit Analyses**

Table 20 provides the range of fit statistics for the survey items from the five random samples. The item fit statistics were used to indicate items that deviated from the expected hierarchical pattern.
<table>
<thead>
<tr>
<th>Survey Items</th>
<th>Infit</th>
<th>Outfit</th>
<th>Point Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were you able to find a nurse or aide when you wanted one?</td>
<td>.75 to .95</td>
<td>.76 to .92</td>
<td>.65 to .71</td>
</tr>
<tr>
<td>Did the nurses and aides treat your family member with courtesy and respect?</td>
<td>.80 to .89</td>
<td>.77 to .88</td>
<td>.60 to .65</td>
</tr>
<tr>
<td>Did the nurses and aides treat your family member with kindness?</td>
<td>.76 to .84</td>
<td>.76 to .79</td>
<td>.64 to .68</td>
</tr>
<tr>
<td>Did you feel that the nurses and aides really cared about your family member?</td>
<td>.79 to .98</td>
<td>.77 to .87</td>
<td>.72 to .76</td>
</tr>
<tr>
<td>Did you ever see any nurses or aides be rude to your family member or another resident?</td>
<td>.84 to 1.04</td>
<td>.69 to .77</td>
<td>.41 to .49</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to eat?</td>
<td>.81 to 1.05</td>
<td>.76 to .88</td>
<td>.46 to .55</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to drink?</td>
<td>.79 to .95</td>
<td>.76 to .89</td>
<td>.50 to .58</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to toilet?</td>
<td>.79 to 1.09</td>
<td>.78 to .95</td>
<td>.49 to .65</td>
</tr>
<tr>
<td>Did your family member look and smell clean?</td>
<td>.92 to 1.01</td>
<td>.98 to 1.05</td>
<td>.61 to .65</td>
</tr>
<tr>
<td>Did nurses and aides handle situations in a way you felt was appropriate?</td>
<td>.73 to 1.07</td>
<td>.74 to .99</td>
<td>.60 to .70</td>
</tr>
<tr>
<td>How often did you get information as soon as you wanted?</td>
<td>.79 to .87</td>
<td>.77 to .85</td>
<td>.68 to .71</td>
</tr>
<tr>
<td>Did the nurses and aides explain things in a way that was easy for you to understand?</td>
<td>.84 to 1.14</td>
<td>.78 to 1.06</td>
<td>.59 to .65</td>
</tr>
</tbody>
</table>
Did the nurses or aides ever discourage you from asking questions?  .86 to 1.01  .78 to .99  .19 to .31

Did your family member’s room look and smell clean?  .80 to .90  .79 to .93  .66 to .71

Did the public areas of the nursing home look and smell clean?  .99 to 1.19  .98 to 1.10  .57 to .61

How often were your family member’s medical belongings lost or damaged?  1.94 to 2.10  2.37 to 2.86  .36 to .39

How often were clothes lost or damaged?  1.39 to 1.44  1.69 to 1.81  .51 to .56

Did you ever stop yourself from talking to staff about your concerns because you thought they may take it out on your family member?  .94 to 1.08  .83 to 1.07  .37 to .51

Were you involved as much as you wanted to be in decisions about your family member’s care?  1.22 to 1.30  1.22 to 1.70  .45 to .55

Using a number from 0 to 10, rate the care at this nursing home. (Collapsed to 6 points)  .76 to .79  .75 to .79  .82 to .85

If someone needed nursing home care, would you recommend this nursing home to them?  .74 to .78  .74 to .79  .72 to .73

Did you feel there were enough nurses and aides?  .97 to 1.14  .97 to 1.26  .71 to .76

Did you get all the information you wanted about payments and expenses?  1.52 to 1.73  1.48 to 1.71  .46 to .57

Infit statistics above 1.3 represent unexpected person responses, which often indicate an item that is measuring a different construct. Three items (How often were your family member’s medical belongings lost or damaged, How often were clothes lost
or damaged and Did you get all the information you wanted about payments and expenses) had infit mean squares above 1.3 criterion (Bond and Fox, 2007) in all of the random samples. These are the same items that misfit when the entire data set was used. Two of these items used the same response scale type (*never, once or twice, and three or more times*). The rating scale introduced noise into the measurement process for these items which may have also led to their infit mean squares above the 1.3 threshold.

The same three items also had outfit mean squares above 1.3. This means that these items indicated idiosyncratic variations from expectation. All of the items had positive point-measure correlations. Positive correlations are an indication that item responses are in the direction of the latent variable. Many of the items had high point-measure correlations (above .50) meaning that higher ratings on these items were related to higher person measures.

**Variable Meaning**

Having established a reliably calibrated measure with items that fit reasonably well within a theoretically-justified continuum, the next step was to interpret the family satisfaction construct. The first step of this process was to examine the measure score, which provides information about the items that were easy and difficult for respondent to provide a favorable response. Table 21 lists the items from easiest to most difficult for family members to provide a favorable response.

The measure scores for the survey items range from -2.74 to 2.17 logits which means the item difficulty covers nearly five logits. Most of the items had negative measure scores which means that respondents found the items easy to provide a favorable response. This could be an indication of a possible targeting issue.
### Table 21

**Range of Item Measures and Standard Errors**

<table>
<thead>
<tr>
<th>Survey Items</th>
<th>Measure</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the nurses or aides ever discourage you from asking questions?</td>
<td>-2.74 to -2.08</td>
<td>.27 to .32</td>
</tr>
<tr>
<td>Did the nurses and aides treat your family member with courtesy and respect?</td>
<td>-1.21 to -1.03</td>
<td>.10 to .11</td>
</tr>
<tr>
<td>Did the nurses and aides treat your family member with kindness?</td>
<td>-1.03 to -.92</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Were you involved as much as you wanted to be in decisions about your family member’s care?</td>
<td>-.95 to -.79</td>
<td>.10 to .11</td>
</tr>
<tr>
<td>Did you get all the information you wanted about payments and expenses?</td>
<td>-.88 to -.64</td>
<td>.12 to .14</td>
</tr>
<tr>
<td>Did the nurses and aides explain things in a way that was easy for you to understand?</td>
<td>-.79 to -.99</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Did the public areas of the nursing home look and smell clean?</td>
<td>-.75 to -.53</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Did nurses and aides handle situations in a way you felt was appropriate?</td>
<td>-.70 to -.28</td>
<td>.15 to .16</td>
</tr>
<tr>
<td>Did you ever see any nurses or aides be rude to your family member or another resident?</td>
<td>-.36 to -.20</td>
<td>.15 to .16</td>
</tr>
<tr>
<td>Did you feel that the nurses and aides really cared about your family member?</td>
<td>-.29 to -.12</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>How often did you get information as soon as you wanted?</td>
<td>-.22 to .03</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Did you ever stop yourself from talking to staff about your concerns because you thought they may take it out on your family member?</td>
<td>-.20 to -.04</td>
<td>.19 to .22</td>
</tr>
<tr>
<td>Were you able to find a nurse or aide when you wanted one?</td>
<td>-.06 to .10</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Question</td>
<td>Score Range</td>
<td>Standard Error</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to eat?</td>
<td>-.06 to .27</td>
<td>.18 to .21</td>
</tr>
<tr>
<td>Did your family member look and smell clean?</td>
<td>.04 to .10</td>
<td>.08 to .09</td>
</tr>
<tr>
<td>Did your family member’s room look and smell clean?</td>
<td>.06 to .35</td>
<td>.09 to .11</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to drink?</td>
<td>.18 to .25</td>
<td>.20 to .22</td>
</tr>
<tr>
<td>How often were your family member’s medical belongings lost or damaged?</td>
<td>.60 to .76</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>If someone needed nursing home care, would you recommend this nursing home to them?</td>
<td>.76 to 1.05</td>
<td>.09 to .10</td>
</tr>
<tr>
<td>Using a number from 0 to 10, rate the care at this nursing home. (Collapsed to 6 points)</td>
<td>.83 to 1.06</td>
<td>.06 to .07</td>
</tr>
<tr>
<td>Nurses or aides either didn’t help or made him or her wait too long to toilet?</td>
<td>1.23 to 1.87</td>
<td>.25 to .27</td>
</tr>
<tr>
<td>Did you feel there were enough nurses and aides?</td>
<td>1.85 to 1.90</td>
<td>.08 to .09</td>
</tr>
<tr>
<td>How often were clothes lost or damaged?</td>
<td>1.96 to 2.17</td>
<td>.10 to .11</td>
</tr>
</tbody>
</table>

The standard error around each measure score provided information about the likelihood two items would switch positions on the item hierarchy. A 95% confidence interval was created by multiplying the error by 1.96 then adding it to one measure score and subtracting it from another measure score to determine whether there was less than a 5% chance these items would change positions in the item hierarchy. Items adjacent or close to each other in the hierarchy had a greater than 5% chance of switching positions. Items further apart (e.g., items 22 and 40) had a less than 5% chance of switching positions in the hierarchy.
Next, the person-item map was examined to assess whether there is a targeting issue with the survey. Figure 7 shows a WINSTEPS person-item map for the family satisfaction variable. The persons span a logit range of approximately 7.5 logits while the items span a logit range of approximately 4.5 logits. This indicates that the survey respondents appear to be more satisfied than the items are able to measure on the survey.

A noticeable gap in the variable definition appears between “How often were your family member’s medical belongings lost or damaged?” and “How often were clothes lost or damaged?” This gap in the variable appears where many of the respondents were located. This could indicate an area where additional items could be created to help further define the variable.

An examination of Figure 7 shows that there may be a targeting issue with the NHCAHPS family survey as the distribution of person abilities is well above the distribution of item difficulties. Even the most difficult items were agreed with by a majority of respondents. A well-targeted survey would have a mean person ability estimate closer to zero.
Figure 7. Person-Item Map. Each # is equal to 3 survey respondents.
To further examine the potential targeting issue and understand the variable meaning, the patient-item map was examined. The general keyform map (see Figure 8) illustrates the ordering of the satisfaction items from least (top) to most (bottom) agreeable, on an equal-interval scale. The “Recommendation to Others” and “Overall rating of Care” items were removed from General Keyform Map for three reasons. First, the “Recommendation to Others” item was not part of the original set of items intended for use in the reporting of the NHCAHPS Family survey results. As discussed previously, this is considered an important item for SNF management teams so it was included in the examination of the survey. Second, these two items are often considered outcome items that are meant to summarize overall perceptions. Finally, the infit mean squares for these items were close to the .75 threshold (see Table 20) meaning that they overfit or were redundant items.

Figure 8 illustrates both the items and respondents in the same, equal-interval, probabilistic frame of reference so that likelihood statements can be made when regarding what would be expected to happen when a certain person encounters a certain item. An informal qualitative analysis was used to categorize the items to obtain a fuller understanding of the nature of the NHCAHPS Family survey’s satisfaction construct.

Figure 8 represents how satisfaction develops. The development of satisfaction begins with items in Group one. Group one included items related to the interpersonal skills of staff members, specifically nurses and aides. These items focused on the general level of respect and openness of nursing staff members. Group two included items related to SNF procedures and communication. These items focused on the processes SNF’s used to maintain the physical environment and communicate with
family members about the care provided. Group three included items related to problem solving and personal interactions with staff members. These items focused on staff being available to provide information or deal with problems and the manner in which staff members responded during interactions. Group four included items related to direct care provided to the patient. These items focused on the amount of help available to patients to attend to biological and environmental needs.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 8. General Keyform Map**

The arrow drawn in Figure 8 through the average for the group (M) indicated that the average family member would most likely respond with an *Always* response for items in groups one and two, between *Usually* and *Always* for items in Group three while items in Group four were the most difficult to agree with. The sample is thus reported
relatively high levels of satisfaction on average. Figure 8 shows that items in Groups one and two provide a description of what the lower end of the scale looks like. These items were not useful in identifying dissatisfied family member across the organization because the organization appears to perform well on this construct. Understanding the lower end of the scale is beneficial for individual SNFs that may be struggling with family satisfaction to identify problem areas.

The NHCAHPS Family Survey for items in Groups three and four as there appears to be a decision to be made by respondents between rating points three and four on most of these items. This also means the NHCAHPS Family Survey appears to be able to measure basic levels of family satisfaction. This means that SNF management teams could use this information to ensure they are meeting basic expectations while SNFs that are performing well on the entire survey may want to construct items to understand higher satisfaction levels beyond what is assessed with items in Group four.

**Principle Contrasts Analysis**

To further examine the dimensionality of the instrument used in this study, a Rasch Principle Contrasts Analysis (RPCA) of Residuals was conducted with each of the random samples of 500 respondents (See Table 22). The results of the RPCA were nearly identical for each of the random samples. Random sample one was chosen to report since it had the lowest variance explained, 71.1%.

The purpose of this analysis was not to construct variables but to explain the amount of variance accounted for the extracted linear measure. According to Table 22, the Rasch dimension explains 71.1% of the variance in the data. The largest secondary dimension, “the first contrast in the residuals” explains 2.7% of the variance, and the
The eigenvalue of the unexplained variance in the first contrast was 2.2, bigger than the strength of two items, the smallest amount that could be considered a “dimension”. However, based on the general rules for RPCA provided by Linacre (2006), the instrument used in this study may be considered as unidimensional, because the variance explained by measures is larger than 60%, the unexplained variance in the first contrast was less than three, and the unexplained variance explained by the first contrast in the residuals was less than 5%, as is shown in Table 22.

Table 22

<table>
<thead>
<tr>
<th>Standardized Residual Variance in Eigenvalue Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations = 79.7 100.0%</td>
</tr>
<tr>
<td>Variance explained by measures = 56.7 71.1%</td>
</tr>
<tr>
<td>Unexplained variance (total) = 23.0 28.9% 100.0%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast = 2.2 2.7% 9.4%</td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast = 1.7 2.0% 7.6%</td>
</tr>
<tr>
<td>Unexplained variance in 3rd contrast = 1.6 2.0% 6.9%</td>
</tr>
<tr>
<td>Unexplained variance in 4th contrast = 1.6 2.0% 6.8%</td>
</tr>
<tr>
<td>Unexplained variance in 5th contrast = 1.5 1.9% 6.4%</td>
</tr>
</tbody>
</table>

Composite Domain Analysis

The above analyses focused on understanding the unidimensionality and variable meaning of the entire NHCAHPS Family Survey. With the intention of CMS to report SNF scores for the four composite domains (Meeting Basic Needs; Nurses/Aides Kindness & Respect towards Resident; Nursing Home Provides Information/Encourages Respondent Involvement; and Nursing Home Staffing, Care of Belongings &
Cleanliness) it was important to conduct the same diagnostic analyses for each composite domain. See Appendix C for a list of the items in each of the composite domains.

**Meeting basic needs composite domain.** The Meeting Basic Needs composite domain includes three items. These items were designed to assess whether family members helped patients with basic needs (i.e., eating) because the nurses or aides didn’t help or made the patient wait too long. Each of these items used a Yes or No rating scale.

The person separation and person reliability were both zero. This means that these three items could not reliably differentiate persons on the measured variable. The item separation was 11.23 and the item reliability was .99. This means that the replicability of item placement along the pathway would be expected to remain consistent if these items were given to a similar group of respondents. With only 3 items, this level item separation and reliability appeared more precise than expected.

The item fit statistics were used to indicate items that deviated from the expected hierarchical pattern. The in-fit and out-fit mean squares were all below the 1.4 threshold indicating that the items in the Meeting Basic Needs composite domain were not capturing something qualitatively different from meeting basic needs. The variance explained by these three items was 57.1%, which was close to the 60% threshold for unidimensionality.

**Nurses and aides kindness and respect towards resident.** The Nurses and Aides Kindness and Respect Towards Resident composite domain includes five items. Four of the items used the Never, Sometimes, Usually, and Always rating scale. The remaining two items used a Yes or No rating scale.
Rasch person and item separation statistics were 1.73 and 15.69, respectively. The person separation of 1.73 translated into less than 3 statistically distinct strata. This means that persons were measured on a dichotomy rather than a continuum. Item separation was 15.69 which translated into 21 distinct strata.

Rasch person reliability was low at .68. This means that the replicability of the person ordering that would be expected if this sample were given a parallel set of items measuring the same construct was low. The item reliability was high at 1.0 meaning the item ordering along the pathway would be expected to remain consistent if these items were given to another sample.

The average measure and the step thresholds for the rating scale categories advanced monotonically from Never to Always. The step threshold increase between each rating scale point was above the 1.4 threshold. The step threshold between Usually and Always was 4.88 logits. This was close to the five logit cutoff which means that the rating scale was close to showing large gaps in the variable. Also, the in-fit and out-fit mean square statistics were above the 2.0 threshold for the Never response category. This is an indication that this rating scale category was introducing noise into the measurement process.

Two items had in-fit and out-fit mean squares above the 1.4 threshold indicating that these items were capturing something qualitatively different from nurses and aides kindness and respect. The variance explained by these five items was 53.1%, which was below the 60% threshold for unidimensionality.
Nursing home provides information and encourages respondent involvement.

This composite domain includes six items. Four of the items used the Never, Sometimes, Usually, and Always rating scale. The fifth item used a Yes or No rating scale.

Rasch person and item separation statistics were 1.07 and 16.44, respectively. The person separation of 1.07 translated into less than two statistically distinct strata. This means that person groups were indistinguishable from one another. Item separation was 16.44 which translated into 21 distinct strata.

Rasch person reliability was low at .44. This means that the replicability of the person ordering that would be expected if this sample were given a parallel set of items measuring the same construct was low. The item reliability was high at 1.0 meaning the item ordering along the pathway would be expected to remain consistent if these items were given to another sample.

The average measure for both rating scale types advanced monotonically. The step thresholds for the 4-point rating scale categories advanced monotonically from Never to Always. The step threshold increase between each rating scale point was above the 1.4 threshold meaning that respondents did reliably distinguish between the rating scale categories. Also, the in-fit and out-fit mean square statistics were all below the 2.0 threshold for both rating scale types.

Two items had in-fit and out-fit mean squares above the 1.4 threshold indicating that these items were capturing something qualitatively different from providing information and encouraging respondent involvement. The variance explained by these six items was 72.1%, which was above the 60% threshold for unidimensionality.
**Nursing home staffing, care of belongings and cleanliness.** This composite domain includes seven items. Five of the items used the *Never, Sometimes, Usually,* and *Always* rating scale. The remaining two items item used a *Never, Once or Twice* and *Three or More Times* rating scale.

Rasch person and item separation statistics were 1.61 and 31.37, respectively. The person separation of 1.61 translated into less than two statistically distinct strata. Item separation was 31.37 which translated into 42 distinct strata.

Rasch person reliability was low at .73. This means that the replicability of the person ordering that would be expected if this sample were given a parallel set of items measuring the same construct was low. The item reliability was high at 1.0 meaning the item ordering along the pathway would be expected to remain consistent if these items were given to another sample.

The average measure for both rating scale types advanced monotonically. The step thresholds for the 4-point rating scale categories advanced monotonically from *Never* to *Always*. The step threshold increase between each rating scale point was above the 1.4 threshold meaning that respondents did reliably distinguish between the rating scale categories. Also, the in-fit and out-fit mean square statistics were all below the 2.0.

The step thresholds for the 3-point rating scale categories advanced monotonically from *Never* to *Three or More Times*. The step threshold increase between each rating scale point was below the 1.4 threshold meaning that respondents did not reliably distinguish between the rating scale categories. Also, the in-fit and out-fit mean square statistics were all below the 2.0 threshold. The variance explained by these seven items was 58.4%, which was slightly below the 60% threshold for unidimensionality.
Chapter Five

Discussion

The purpose of this study was to construct a measure from the NHCAHPS Family Survey using Rasch analysis. Given the critical implications (e.g., comparisons of SNFs on NHCAHPS Family Survey scores) and decisions resulting from performance on the NHCAHPS Family Survey (i.e., Medicare reimbursement) it was imperative to construct a measure of family satisfaction.

Published research (e.g., Sangle et al., 2007) focused on the NHCAHPS Resident Survey not the NHCAHPS Family Survey. The only published research using the NHCAHPS Family Survey was reported in the American Institute of Research (AIR) Final Report (2008). The AIR report focused on the development of the four composite domains rather than an assessment of whether the NHCAHPS Family Survey can be used to construct a linear measure of family satisfaction.

Rasch analyses provide several unique contributions to understanding the functionality of the NHCAHPS Family Survey. One of the fundamental goals of Rasch analysis is the development of clear, functional, linear variables. Rasch analysis evaluates the clarity of the construct being measured by the survey, through both summary statistics (separation and reliability) and the person-item map. Because person and item parameters are estimated separately, Rasch estimates are considered sample-independent indicators of how well the survey is able to measure the construct. Traditional parametric statistical models, by comparison, do not provide inferential information beyond the sample studied when used on ordinal raw data. This is a commonly overlooked limitation.
This study focused exclusively on Rasch analysis for assessing the measurement properties of the NHCAHPS Family Survey. Rasch analyses include multiple detailed diagnostics, (including several fit and point-biserial statistics) on both individual item and total survey levels. These diagnostics are used to assess the functionality of items working together as a single dimension and may also be used to evaluate the possible existence secondary constructs, with RPCA.

**Summary of Rasch Findings.**

The fundamental Rasch analyses suggest that the NHCAHPS Family Survey functioned well. The items used to measure the family satisfaction were reliable, and produced low measurement error resulting in high person and item reliability and separation.

Three of the five rating scales were functional, because the scales monotonically progressed along the linear relationship. Respondents were unable to reliably distinguish between response categories with the 4-point rating scale used for the Recommendation to Others item. Results indicated that respondents were unable to distinguish between category one (*definitely no*) and two (*probably no*). These two categories were collapsed together because collapsing these response categories made sense qualitatively (Wright, 1987). This means that collapsing these categories together resulted in a category that represented “No”. By collapsing these categories together, the rating scale for the Recommendation to Others item became a 3-point scale (*no, probably yes, and definitely yes*).

The other rating scale that was not able to reliably distinguish between response categories was the 11-point scale used by the Overall rating of Care item. Results
indicated that only six of the rating scale points could be reliably distinguished by respondents. Research suggests that, although the addition of response categories generally increases reliability, it does so only if the additional categories are not arbitrary (Linacre, 1995). Due to the focus many family members put on this item, additional improvement recommendations to the rating scale are discussed below.

Results from the Principle Components Analysis of Residuals (RPCA) indicated that the NHCAHPS Family Survey was a unidimensional measure of family satisfaction. However, results indicated that there were not enough items at the end of the measure to reliably measure the most satisfied family members. As seen in Figure 7, the person spanned a logit range of approximately 7.5 logits while the items spanned a logit range of approximately 4.5 logits. These results indicated a possible targeting issue. The typical solution to targeting issues is to add additional items to the survey that would span a larger logit range and better match survey respondents.

The targeting issue may be less critical for the NHCAPS Family Survey. Ejaz et. al. (2003) claim that a large scale satisfaction surveys used to compare health care facilities needs to provide consumers with basic information about the quality of staff, the physical environment, and care provided at a facility. If a facility wants to exceed these basic expectations for consumers then they need to assess and monitor these areas independently. This means that a SNF that wants to assess whether its consumers report more than a basic level of satisfaction with the facility and staff need to collect this information themselves. Figure 7 shows that the NHCAHPS Family Survey was able to measure the basic levels of family satisfaction.
In addition to examining the functioning of the entire NHCAHPS Family Survey, the four composite domains reported in the AIR report (2008) were also tested. There were several Rasch diagnostic indices that revealed all of the composite domains were not functioning well. First, the person separation was less than 2.0 for all of the composite domains. This means that person groups were indistinguishable from one another. Second, three of the four composite domains had less than 60% of the variance explained as shown in the RPCA. Third, the Rasch person reliability was low for all of the composite domains. Finally, several of the fit statistics for items within the domains were above the 1.4 threshold indicating these items were capturing something qualitatively different than the intended construct of the composite domain.

**Application of Findings**

The items on the survey separated the respondents into four distinct groups (see Figure 8). The four distinct groups of items were labeled as staff interpersonal skills, SNF procedures, staff problem solving/interactions, and direct care provided by staff members. Understanding these four groups of items and how the items increase in difficulty is a tool that will assist SNF management teams to intervene directly with individual survey respondents based on their individual responses and in their development of facility-wide improvement plans.

The general keyform map provides information on whether the response to an item is higher or lower than expected. SNF management teams could be provided with the response pattern of individual respondents of respondents with negative measure scores (most dissatisfied) to identify the items these respondents rated the facility higher or lower than expected. Having access to this information would provide SNF
management teams into areas they need to contact the respondent about as part of a service recovery process. Service recovery would help improve the SNF’s growth and reputation because the SNF management team could intervene with individual family members. Reicheld (2003) reported that addressing family concerns reported on satisfaction surveys leads to increased satisfaction and recommendation of service to others.

Figure 9 provides an example of the responses provided by an actual survey respondent to illustrate how this information could be used by SNF management teams. This family member provided responses that resulted in Rasch fit indices (i.e., an infit mean square of 2.18, outfit mean square of 3.26 and point-measure correlation of .10) categorizing him as misfitting. This means that the indices that estimated the extent to which his responses conformed to the Rasch model were beyond expectations. This family member also had an measure score of -1.07 logits. This means that this family member found it difficult to agree with many of the survey items.

The arrow in Figure 9 represents this family member’s measure score (-1.07 logits). The arrow indicates the expected survey responses for a family member with this measure score. The rectangles next to each survey item are the actual responses provided by this family member. The rectangles extend between the threshold points from the lower response category to the threshold point of the upper response category. The threshold points indicate where there is a 50% chance a respondent would choose between each of the response categories.

An examination of Figure 9 shows several trends from the family member’s responses that would be beneficial for intervening with this person on areas where this
person scored lower than expected given his measure score. First, responses to item 13 and item 3 were lower than expected. These items are part of the “Interpersonal Skills of Staff Members” group of items. These items focus on the interpersonal skills of nurses and aides. In addition, the response to item 10 was also below expectations. This item also related to how nurses and aides handled situations with this family member. This information could be used by the Director of Nursing to find out more specific information or incidents that occurred when the family visited which resulted in these ratings.

---

**Figure 9. An Individual Survey Response for Service Recovery Intervention**

Figure 9 also displays items that the family member rated higher than expected given his measure score. This family member provided the highest possible rating for items 14 and 15 when he was expected to provide a low rating. Access to this
information would allow the Environmental Services Manager to follow-up with this family member to identity reasons why this person rated these items so high to ensure this level of service continues in the future.

A second application of this study’s findings is changing the way scores are reported to SNFs. Currently, SNF management teams examine two sources of information, the key drivers table and the lowest scoring item, when determining areas for improvement (DerGurhian, 2009). The key drivers table rank orders the survey items based on their correlation with the Overall rating of Care item. This means that SNF management teams may be making decisions on areas for improvement based on statistical analyses while disregarding the underlying meaning of the ratings.

If the goal is to achieve family satisfaction, then the SNF management team needs to examine their performance on items related to the interpersonal aspect of staff members first. Scores on these items could be higher relative to other items but may be lower than expected as indicated in Figure 8. Therefore, determination of areas for improvement needs to start at with interpersonal aspects of staff members first and move up to the other groups of items.

The focus on improving performance on items related to the direct care provided to the patient (Group four) is evident in the improvement plans submitted by many SNF teams. The focus of many of these items is on timeliness of response but, as seen in Figure 8, the improvement of family satisfaction begins with how they feel they are treated first and then timeliness of response second.

A third application is using this information to expand the definition of quality in SNFs and move towards a “consumerist” approach (Berwick, 2009). In a “consumerist”
approach, the patient and family member determines whether quality standards are being met at the health care facility rather than the staff or a third-party governing body (i.e., CMS). The “consumerist” approach could be met through the utilization of this study’s findings to begin developing a Five-Star Quality rating based on customer satisfaction survey results.

Currently, the Five-Star Quality rating system does not include an indicator of customer satisfaction. There is an intention to utilize the NHCAHPS results to either develop a separate Five-Star rating or combine them with one of the current ratings (e.g., Quality Measures). There are three separate NHCAHPS surveys (Family, Resident and Short-Stay Patient) that would be collected as part of the Five-Star rating.

The current study only constructed a measure from the NHCAHPS Family Survey. Similar studies would need to be conducted with the remaining NHCAHPS surveys and then a process could be developed to calibrate these scores together as a way to report customer satisfaction in the same format used by the other Five-Star Quality ratings.

**Improvement Recommendations.**

Despite the Rasch analyses indicating that the NHCAHPS Family Survey functioned well there are still some potential improvements that can be made to improve the NHCAHPS Family Survey. Due to the importance of the Overall rating of Care item, many of the improvement recommendations are focused on this item.

**General changes.** There are currently 50 items on the NHCAHPS Family Survey but only 22 of them are set to be publicly reported on the Nursing Home Compare website. This means that more than half of the collected survey items are not used for a
specific reporting purpose. Some of these items could either be eliminated or changed to allow them to be used as part of the measure.

Elimination of some of the non-reported survey items could help improve survey response rate. Dillman (1991) reported survey length negatively impacts response rate. Reducing the number of items could result in respondents perceiving the survey as a less arduous and time consuming task thus increasing the likelihood of response. Increasing response rate decreases the margin of error around an individual SNF’s survey scores.

A second option for some of the non-reported items would be to change them with the purpose of adding them the measure. Of course, the addition of any survey item would necessitate a reexamination of the Rasch diagnostics to assess fit and potential changes in the item ordering that could impact variable meaning.

Changing some of the non-reported items could address the potential targeting issue noted above. Items could be written with the intention to increase the logit range covered by the items. In addition, reducing unnecessary items could allow for more items to be added to the composite domains. Composite domains are currently used in every other version of the CAHPS surveys (e.g., HHCAHPS). Therefore, CMS may want to continue this practice so adding appropriate items to these composite domains could help improve their functioning as a measure.

**Overall rating of care.** The Overall rating of Care item is the only item on the survey that does not contain written descriptors at each response category. There are only descriptors provided at the 0 (*worst possible care*) and 10 (*best possible care*) response categories. Research is needed to provide a descriptor at each of the points on the revised 6-point scale. This could be done through a qualitative analysis of the written
comments provided by survey respondents based on their response to the Overall rating of Care item. Focus groups with family members could also be conducted to provide descriptions of care at each response category. Following the development of rating category descriptors, the item should be retested to assess the Rasch rating scale diagnostics.

The need for better descriptors at each rating scale point was also evident when no respondents used the zero (worst possible care) rating. It is hypothesized that this response category was not used because family members would not want to believe they placed a resident in a SNF that provides the worst possible care. An alternative hypothesis is that many family members do not have experience with another SNF making them unaware of the level of care provided at another SNF.

**Limitations.**

A key limitation of this study is the fact that the family members who completed this survey were all from SNFs within one organization. It will be important to validate these findings with family members of residents residing at SNFs in other organizations. The SNFs used in this study could be unique in terms of the processes implemented that impact satisfaction reported by family members.
References


Castle, N. G. and Enberg, J. (2005). Staff Turnover and Quality of Care in Nursing Homes. *Medical Care, 43*(6), 616 – 626.


IBISWorld (2012). Coming of Age: An Older Population will Spur Demand, but Reimbursements will shrink. IBISWORLD Industry Report 62311 Nursing Care Facilities in the US.


Mor, V., Berg, K., Angelelli, J., Gifford, D., Morris, J., and Moore, T., (2003). The


Schwartz, M., Cramer, I. E., Holmes, S. K., Cohen, A. B., Restuccia, J. D., Lukas, C.,


Appendix A

NCAHPS Family Survey

The Resident

1. Who is the person listed in the cover letter?
   1. My Spouse/Partner
   2. My Parent
   3. My Mother-in-Law/Father-in-Law
   4. My Grandparent
   5. My Aunt or Uncle
   6. My Sister or Brother
   7. My Child
   8. My Friend
   9. Other
   Please print: _____________________

2. For this survey, the phrase “family member” refers to the person listed in the cover letter. Is your family member now living in the nursing home listed in the cover letter?
   1. Yes → If Yes, go to #4
   2. No

3. Was your family member discharged from this facility or did he or she die?
   1. Discharged → If Discharged, Stop Here. Please return this survey in the postage-paid envelope.
   2. Deceased → If your family member is deceased, we understand that you may not want to fill out a survey at this time.

   If you would like to fill out the rest of the survey, we would be very grateful for your feedback. Please go to #5 and answer the questions about your family member’s last 6 months at the nursing home. Thank you for your help.

4. Do you expect your family member to live in this or any other nursing home permanently?
   1. Yes
   2. No
   3. Don’t know

5. In total, about how long has your family member lived in this nursing home?
   1. Less than 1 month → If less than 1 month, Stop Here. Please return this survey in the postage-paid envelope.
   2. 1 month to almost 3 months
   3. 3 months to almost 6 months
   4. 6 months to almost 12 months
   5. 12 months or longer

6. In the last 6 months, has your family member ever shared a room with another person at this nursing home?
   1. Yes
   2. No

7. Does your family member have serious memory problems because of Alzheimer’s disease, dementia, a stroke, an accident, or something else?
   1. Yes
   2. No

8. In the last 6 months, how often was your family member capable of making decisions about his or her own daily life, such as when to get up, what clothes to wear, and which activities to do?
   1. Never
   2. Sometimes
   3. Usually
   4. Always
Your Visits

Please answer the following questions only for yourself. Do not include the experiences of other family members.

9. In the last 6 months, about how many times did you visit your family member in the nursing home?
   1. 0 to 1 time → If 0 to 1, go to #43 on page 6
   2. 2 to 5 times
   3. 6 to 10 times
   4. 11 to 20 times
   5. More than 20 times

10. In the last 6 months, during any of your visits, did you try to find a nurse or aide for any reason?
    1. Yes
    2. No → If No, go to #12

11. In the last 6 months, how often were you able to find a nurse or aide when you wanted one?
    1. Never
    2. Sometimes
    3. Usually
    4. Always

12. In the last 6 months, how often did you see the nurses and aides treat your family member with courtesy and respect?
    1. Never
    2. Sometimes
    3. Usually
    4. Always

13. In the last 6 months, how often did you see the nurses and aides treat your family member with kindness?
    1. Never
    2. Sometimes
    3. Usually
    4. Always

14. In the last 6 months, how often did you feel that the nurses and aides really cared about your family member?
    1. Never
    2. Sometimes
    3. Usually
    4. Always

15. In the last 6 months, did you ever see any nurses or aides be rude to your family member or any other resident?
    1. Yes
    2. No

16. In the last 6 months, during any of your visits, did you help your family member with eating?
    1. Yes
    2. No → If No, go to #18

17. Was it because the nurses or aides either didn’t help or made him or her wait too long?
    1. Yes
    2. No
18. In the last 6 months, during any of your visits, did you help your family member with drinking?

☐ Yes
☐ No → If No, go to #20

19. Was it because the nurses or aides either didn’t help or made him or her wait too long?

☐ Yes
☐ No

20. 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?

☐ Yes
☐ No → If No, go to #22

21. Was it because the nurses or aides either didn’t help or made him or her wait too long?

☐ Yes
☐ No

22. In the last 6 months, how often did your family member look and smell clean?

☐ Never
☐ Sometimes
☐ Usually
☐ Always

23. 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?

☐ Yes
☐ No → If No, go to #25

24. In the last 6 months, how often did the nurses and aides handle the situation in a way that you felt was appropriate?

☐ Never
☐ Sometimes
☐ Usually
☐ Always

Your Experience With Nurses & Aides

25. In the last 6 months, did you want to get information about your family member from a nurse or an aide?

☐ Yes
☐ No → If No, go to #27

26. In the last 6 months, how often did you get this information as soon as you wanted?

☐ Never
☐ Sometimes
☐ Usually
☐ Always
27. In the last 6 months, how often did the nurses and aides explain things in a way that was easy for you to understand?
   1. Never
   2. Sometimes
   3. Usually
   4. Always

28. In the last 6 months, did the nurses and aides ever try to discourage you from asking questions about your family member?
   1. Yes
   2. No

29. In the last 6 months, how often did your family member’s room look and smell clean?
   1. Never
   2. Sometimes
   3. Usually
   4. Always

30. In the last 6 months, how often did the public areas of the nursing home look and smell clean?
   1. Never
   2. Sometimes
   3. Usually
   4. Always

31. Personal medical belongings are things like hearing aids, glasses, and dentures. In the last 6 months, how often were your family member’s personal medical belongings damaged or lost?
   1. Never
   2. Once
   3. Two or more times

32. In the last 6 months, did your family member use the nursing home’s laundry service for his or her clothes?
   1. Yes
   2. No → If No, go to #34

33. In the last 6 months, when your family member used the laundry service, how often were clothes damaged or lost?
   1. Never
   2. Once or twice
   3. Three or more times

34. In the last 6 months, were you ever unhappy with the care your family member received at the nursing home?
   1. Yes
   2. No → If No, go to #36

35. 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?
   1. Yes
   2. No
39. If someone needed nursing home care, would you recommend this nursing home to them?
   1. Definitely no
   2. Probably no
   3. Probably yes
   4. Definitely yes

40. In the last 6 months, how often did you feel there were enough nurses and aides in this nursing home?
   1. Never
   2. Sometimes
   3. Usually
   4. Always

You And Your Role

Please remember that the questions in this survey are about your experiences. Do not include the experiences of other family members.

41. In the last 6 months, did you ask the nursing home for information about payments or expenses?
   1. Yes
   2. No → If No, go to #43

42. In the last 6 months, how often did you get all the information you wanted from the nursing home about payments or expenses?
   1. Never
   2. Sometimes
   3. Usually
   4. Always

Care Of Your Family Member

36. In the last 6 months, have you been involved in decisions about your family member’s care?
   - 1. Yes
   - 2. No → If No, go to #38

37. 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?
   - 1. Never
   - 2. Sometimes
   - 3. Usually
   - 4. Always

Overall

38. 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. Worst care possible
   - 1
   - 2
   - 3
   - 4
   - 5
   - 6
   - 7
   - 8
   - 9
   - 10. Best care possible
About You

43. What is your age?
   1. 18 to 24
   2. 25 to 34
   3. 35 to 44
   4. 45 to 54
   5. 55 to 64
   6. 65 to 74
   7. 75 or older

44. Are you male or female?
   1. Male
   2. Female

45. What is the highest grade or level of school that you have completed?
   1. 8th grade or less
   2. Some high school but did not graduate
   3. High school graduate or GED
   4. Some college or 2-year degree
   5. 4-year college graduate
   6. More than 4-year college degree

46. Are you of Hispanic or Latino origin or descent?
   1. Yes, Hispanic or Latino
   2. No, not Hispanic or Latino

47. What is your race? Mark one or more.
   1. White
   2. Black or African-American
   3. Asian
   4. Native Hawaiian or other Pacific Islander
   5. American Indian or Alaska Native
   6. Other

48. What language do you mainly speak at home?
   1. English
   2. Spanish
   3. English and Spanish equally
   4. Some other language

49. Did someone help you complete this survey?
   1. Yes
   2. No → If No, go to #51

50. How did that person help you? Mark one or more.
   1. Read the questions to me
   2. Wrote down the answers I gave
   3. Answered the questions for me
   4. Translated the questions into my language
   5. Helped in some other way

Please print: __________________________

______________________________

______________________________
Optional

51. Is there anything else you'd like to say about the care your family member gets at this nursing home?

*Please print:*

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

Thank you.

Please return the completed survey in the postage-paid envelope.
Appendix B

Composite Measure Proportional Scoring Method Description and Example

The calculation of NHCAHPS survey composites uses a proportional scoring method, which basically generate a proportion for each response option. There are two basic steps to this approach:

1. Calculate the proportion of responses in each response category for each item in a composite.

2. Combine these proportions for all items in a composite.

This approach allows for the reporting of “top box” scores. “Top box” scoring involves reporting only the score for the most positive categories (e.g., the proportion of respondents reporting “Always”).

The example below provides greater detail of the two steps in the proportional scoring method.

Example

Given a composite measure with five items, where each item has four response options, a facility’s score for that composite is the proportion of responses (excluding missing data) in each response category. The following steps show how those proportions are calculated:

Step 1 – Calculate the proportion of cases in each response category for the first question:

P11 = Proportion of respondent who answered “never”

P12 = Proportion of respondent who answered “sometimes”
P13 = Proportion of respondent who answered “usually”
P14 = Proportion of respondent who answered “always”

Follow the same steps for the second question:
P21 = Proportion of respondent who answered “never”
P22 = Proportion of respondent who answered “sometimes”
P23 = Proportion of respondent who answered “usually”
P24 = Proportion of respondent who answered “always”

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

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:

PC1 = Composite proportion who said “never” = (P11+P21+P31+P41+P51)/5
PC2 = Composite proportion who said “sometimes” = (P12+P22+P32+P42+P52)/5
PC3 = Composite proportion who said “usually” = (P13+P23+P33+P43+P53)/5
PC1 = Composite proportion who said “always” = (P14+P24+P34+P44+P54)/5
Appendix C

Composite Measures for the NHCAHPS Family Survey

**Meeting Basic Needs: Help with Eating, Drinking and Toileting**

Q17 Wait too long for help with eating  
Response Options
● Yes
● No

Q19 Wait too long for help with drinking

Q21 Wait too long for help with toileting

**Nurses/Aides’ Kindness/Respect Towards Resident**

Q12 Nurses/Aides treat resident with courtesy and respect  
Response Options
● Never
● Sometimes
● Usually
● Always

Q13 Nurses/Aides treat resident with kindness

Q14 Nurses/Aides really cared about the resident

Q24 Nurses/Aides appropriate with resident displaying behavioral problems

Q15 Nurses/Aides rude to resident  
Response Options
● Yes
● No

**Nursing Home Provides Information/Encourages Respondent Involvement**

Q26 Nurses/Aides give respondent timely information about resident  
Response Options
● Never
● Sometimes
● Usually
● Always

Q27 Nurses/Aides explain things to respondent

Q37 Respondent involved in decisions about care

Q42 Respondent given information about payments/expenses

Q28 Nurses/Aides discourage respondents questions  
Response Options
● Yes
● No

Q35 Respondent stops self from complaining
**Nursing Home Staffing, Care of Belongings, and Cleanliness**

<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q11</td>
<td>Can find a nurse or aide</td>
<td>Response Options</td>
</tr>
<tr>
<td>Q40</td>
<td>Enough nurses/aides</td>
<td>● Never</td>
</tr>
<tr>
<td>Q29</td>
<td>Room looks/smells clean</td>
<td>● Sometimes</td>
</tr>
<tr>
<td>Q22</td>
<td>Resident looks/smells clean</td>
<td>● Usually</td>
</tr>
<tr>
<td>Q30</td>
<td>Public areas look/smell clean</td>
<td>● Always</td>
</tr>
<tr>
<td>Q31</td>
<td>Family member’s personal medical belongings lost</td>
<td>Response Options</td>
</tr>
<tr>
<td>Q33</td>
<td>Family member’s clothes damaged or lost</td>
<td>● Never</td>
</tr>
</tbody>
</table>

- Q29: Room looks/smells clean
- Q22: Resident looks/smells clean
- Q30: Public areas look/smell clean
- Q31: Family member’s personal medical belongings lost
- Q33: Family member’s clothes damaged or lost
Appendix D

Detailed Description of Scale Collapsing for Overall Quality of Care Item

Appendix D

Collapsing Option A.

In Option A, a more uniform distribution of response categories was created. Respondents chose categories one, two and three 28, 25 and 56 times respectively. Thus, these three response categories were collapsed together resulting in 109 total respondents using this collapsed category making it the least frequently chosen. Response categories four and five were collapsed together resulting in 331 total respondents. Response categories six and seven were collapsed together resulting in 809 total respondents. Response options 8, 9 and 10 which had 1177, 1082 and 978 responses respectively were not collapsed together. Collapsing Option A resulted in six response categories. Table 1 shows the rating scale diagnostics for Option A.

Table 1

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>109</td>
<td>-.93</td>
<td>.71</td>
<td>.73</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>331</td>
<td>-.08</td>
<td>.65</td>
<td>.65</td>
<td>-2.36</td>
<td>.11</td>
</tr>
<tr>
<td>Three</td>
<td>809</td>
<td>.85</td>
<td>.66</td>
<td>.63</td>
<td>-1.35</td>
<td>.06</td>
</tr>
<tr>
<td>Four</td>
<td>1177</td>
<td>1.83</td>
<td>.70</td>
<td>.65</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>Five</td>
<td>1082</td>
<td>2.78</td>
<td>.76</td>
<td>.76</td>
<td>1.35</td>
<td>.04</td>
</tr>
<tr>
<td>Six</td>
<td>978</td>
<td>3.87</td>
<td>.82</td>
<td>.86</td>
<td>2.36</td>
<td>.05</td>
</tr>
</tbody>
</table>
The average measure and the step thresholds for the rating scale categories advanced monotonically from rating scale categories one to six. Compared to the original rating scale categories, the increases were more uniform (i.e., close to 1 logit). Also, the step threshold increases between ratings scale categories three and four and four and five were both over 1.3 logits but were still below the recommended increase of 1.4. The out-fit and in-fit mean squares were all below 1.0.

Figure 1. Overall Rating of Care Rating Scale Collapsing Option A Probability Curves

Figure 1 provides the probability curves for each rating scale category used in Option A. Examination of the probability curves revealed that each response category did have a distinct peak in the probability curve graph. This means that each rating scale
category was the most probable response for some portion of the measured variable. Also, each of the rating scale categories had close to a .5 probability of response and appeared to be the most probable response across an equal section of the variable. Information found in Table 1 and Figure 1 indicates that respondents were better able to distinguish between response categories in Option A than with the original scale.

**Collapsing Option B.**

Collapsing Option B also created a 6-point rating scale similar to Option A. Option B was attempted based on a visual inspection of the original probability curves. The issues with the rating scale appeared to be centered on response categories 2, 3, 4, 6, and 7. The other response categories (1, 5, 8, 9 and 10) had distinct peaks and spanned a distinct portion of the variable. In Option B, response categories 1, 9 and 10 did not get collapsed with other categories. Response options 2, 3 and 4 were collapsed together. Response options five and six were collapsed. Response options seven and eight were collapsed together. Table 2 shows the rating scale diagnostics for Option B.

The average measure and the step thresholds for the rating scale categories advanced monotonically from rating scale categories one to six. However, as the step thresholds estimates in this table show, the two adjacent categories three and four and categories five and six were not statistically different, less than 1.4 logits apart. This indicated that respondents did not reliably distinguish between these categories. The infit mean square for response category one was 1.0.
Table 2

*Overall Rating of Care Rating Scale Collapsing Option B Diagnostics*

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>28</td>
<td>-1.10</td>
<td>1.00</td>
<td>.98</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>166</td>
<td>-.66</td>
<td>.68</td>
<td>.64</td>
<td>-2.95</td>
<td>.20</td>
</tr>
<tr>
<td>Three</td>
<td>491</td>
<td>.28</td>
<td>.61</td>
<td>.61</td>
<td>-1.43</td>
<td>.09</td>
</tr>
<tr>
<td>Four</td>
<td>1741</td>
<td>1.57</td>
<td>.69</td>
<td>.65</td>
<td>-.64</td>
<td>.05</td>
</tr>
<tr>
<td>Five</td>
<td>1082</td>
<td>2.71</td>
<td>.71</td>
<td>.73</td>
<td>2.15</td>
<td>.04</td>
</tr>
<tr>
<td>Six</td>
<td>978</td>
<td>3.88</td>
<td>.83</td>
<td>.83</td>
<td>2.88</td>
<td>.05</td>
</tr>
</tbody>
</table>

Figure 2 provides the probability curves for each rating scale category used in Option B. Examination of the probability curves revealed that highest probability for endorsement of category three was approximately .4. Information found in Table 2 and from a visual inspection of the probability curves indicated that respondents did not reliably distinguish between response categories three and four. Respondents were unable to make a distinction between response categories two and three and hence used them in an inconsistent manner. Similarly, respondents were unable to distinguish between response categories five and six.
Figure 2. Overall Rating of Care Rating Scale Collapsing Option B Probability Curves

Collapsing Option C.

In Option C, the response categories 1, 2, 3, and 4, responses with a negative average measure were collapsed together. These response categories were also the most redundant visually. Response categories five and six, responses with an average measure between zero and one, were collapsed together. Response categories seven and eight, responses with an average measure between one and two, were collapsed together. Response categories 9 and 10 remained intact. Collapsing Option C resulted in five response categories. Table 3 shows the rating scale diagnostics for Option C.

The average measure and the step thresholds for the rating scale categories advanced monotonically from rating scale categories one to five. The step threshold increase from response category three to response category four was the only increase
that met the guideline of a 1.4 logit increase. The other step threshold increases were less than 1.0 and less than the threshold increases in collapsing Option A. This means that in collapsing Option C, respondents did not reliably distinguish between these categories. The out-fit and in-fit mean squares were all below 1.0.

Table 3

*Overall Rating of Care Rating Scale Collapsing Option C Diagnostics*

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>194</td>
<td>- .73</td>
<td>.69</td>
<td>.99</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>491</td>
<td>.28</td>
<td>.61</td>
<td>.61</td>
<td>-2.07</td>
<td>.09</td>
</tr>
<tr>
<td>Three</td>
<td>1741</td>
<td>1.56</td>
<td>.69</td>
<td>.65</td>
<td>-1.42</td>
<td>.05</td>
</tr>
<tr>
<td>Four</td>
<td>1082</td>
<td>2.76</td>
<td>.71</td>
<td>.72</td>
<td>1.38</td>
<td>.04</td>
</tr>
<tr>
<td>Five</td>
<td>978</td>
<td>3.87</td>
<td>.83</td>
<td>.83</td>
<td>2.11</td>
<td>.05</td>
</tr>
</tbody>
</table>

Figure 3 provides the probability curves for each rating scale category used in Option C. Examination of the probability curves revealed that each response category did have a distinct peak in the probability curve graph. The probability of response for categories two and four was close to .4. The midscale point (response category 3), had approximately a .7 probability of response. Response category three was also the most probable response across a large section of the variable (from about -1.4 to 1.1) meaning that most respondents would have chosen he middle response category in collapsing Option C.
Figure 3. Overall Rating of Care Rating Scale Collapsing Option C Probability Curves

Collapsing Option D.

Collapsing Option D is similar to Option C. The only change in Option D is that response categories 9 and 10 were collapsed together. Visual inspection of the original probability curves showed that response category nine only spanned a small portion of the variable meaning that it would make sense to examine the functioning of the scale with response categories 9 and 10 combined. Collapsing Option D resulted in four response categories. Table 4 shows the rating scale diagnostics for Option D.
Table 4

Overall Rating of Care Rating Scale Collapsing Option C Diagnostics

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Count</th>
<th>Average Measure</th>
<th>In-fit Mean Square</th>
<th>Out-fit Mean Square</th>
<th>Step Threshold</th>
<th>Step Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>194</td>
<td>-.79</td>
<td>.69</td>
<td>.65</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>491</td>
<td>.31</td>
<td>.64</td>
<td>.55</td>
<td>-1.26</td>
<td>.09</td>
</tr>
<tr>
<td>Three</td>
<td>1741</td>
<td>1.62</td>
<td>.66</td>
<td>.57</td>
<td>-.56</td>
<td>.05</td>
</tr>
<tr>
<td>Four</td>
<td>2030</td>
<td>3.29</td>
<td>.68</td>
<td>.78</td>
<td>1.82</td>
<td>.04</td>
</tr>
</tbody>
</table>

The average measure and the step thresholds for the rating scale categories advanced monotonically from rating scale categories one to four. The step threshold increase from response category three to response category four was the only increase that met the guideline of a 1.4 logit increase. The other step threshold increases (response category 2 to 3) was less than 1.0 indicating that respondents did not reliably distinguish between these categories. The out-fit and in-fit mean squares were all below 1.0.

Figure 4 provides the probability curves for each rating scale category used in Option D. Examination of the probability curves revealed that each response category did have a distinct peak in the probability curve graph. The probability of response for category two was close to .4. Response category three had approximately a .6 probability of response and was the most probable response across a smaller section of the variable compared to Option C.
Figure 5. Overall Rating of Care Rating Scale Collapsing Option C Probability Curves