Nursing Home Quality Indicators: Their Uses and Limitations

By

Barbara Manard, Ph.D.
The question of how to measure and improve the quality of care in nursing homes has challenged policymakers for decades. In 1989, researchers at the Center for Health Systems Research and Analysis (CHSRA) at the University of Wisconsin began to address that question by identifying quality indicators (QI) based on the information federal law requires nursing homes to collect about each nursing home resident. That information, known as the Minimum Data Set (MDS), forms the basis of each resident’s care plan, and, as the CHSRA researchers determined, can also be converted into measures that reveal how well a nursing home is caring for its residents. The CHSRA QI consist of specific markers, including the prevalence of pressure ulcers and the use of physical restraints on residents.

Today, QI are being developed and implemented for a broad range of purposes in nursing homes. The federal government and many states plan to rely heavily on QI as part of newer systems for monitoring and regulating nursing home quality. The public disclosure of selected “quality measures” marks the latest use of nursing home QI, and in November 2002, the Centers for Medicare and Medicaid Services (CMS) released 10 quality measures to rate and compare nursing homes in every state. This paper describes the development of the federally sponsored nursing home QI, including the recent development and release of the national quality measures.

As the public starts to rely on disclosure of specific data about nursing homes in choosing a nursing home and regulators use QI to identify poor quality care, it is important to understand the potentially promising uses and substantial limitations of QI. To that end, this paper was commissioned by the Independent Living and Long-Term Care Team of the AARP Public Policy Institute, and was prepared by Barbara Manard, Ph.D. of the Manard Company, an expert in long-term care policy. This report should contribute to understanding the relevant issues.

Faith Mullen, J.D.
Senior Policy Advisor
AARP Public Policy Institute
# TABLE OF CONTENTS

## EXECUTIVE SUMMARY

1

## INTRODUCTION AND BACKGROUND

4

1.0 Introduction ................................................................. 4
1.1 Methodology ................................................................. 4
1.2 Quality Versus Quality Indicators ......................................... 4
1.3 Development of Federally Sponsored QI, Based on Uniform Resident Assessments .......................... 9

## FEDERALLY SPONSORED NURSING HOME QUALITY INDICATORS:
The First Generation ............................................................. 12

2.0 Introduction ........................................................................ 12
2.1 Overview of the 24-Item CHSRA QI ........................................ 12
2.2 Assessing How Well the CHSRA QI Measure Quality or Its Components ............................ 14

2.2.1 Reliability ....................................................................... 15
   - Reliability of the Underlying MDS Data ................................. 15
   - Reliability of the CHSRA QI Per Se .......................................... 16

2.2.2 Validity .......................................................................... 17
   - Face Validity and Content Validity ........................................ 17
   - Testing Against a “Gold Standard” ......................................... 18

2.2.3 New Research on the CHSRA QI ........................................ 20
2.2.4 Additional Issues in Drawing Inferences about Comparative Quality from QI ............................ 21
   - Appropriate Risk Adjustment ............................................. 21
   - Facility Discharge Practices and Short-stay Residents ............... 22
   - Smaller Facilities and Rarer Outcomes ................................ 23

## PUBLIC REPORTING OF QUALITY MEASURES AND OTHER RECENT DEVELOPMENTS

24

3.0 Introduction ........................................................................ 24
3.1 Development Process .......................................................... 24
3.2 Key Issues .......................................................................... 28

3.2.1 The Scope of the Measures ............................................. 28
3.2.2 Risk Adjustment ............................................................ 28

3.2.3 Validity: To What Extent Do CMS’ New Measures Reflect Actual Differences in Nursing Facility Quality of Care? ........................................... 30

## CONCLUSION

32
EXECUTIVE SUMMARY

BACKGROUND

Quality indicators (QI) are markers of potentially poor or excellent health care quality. They may include such elements as mortality rates, avoidable complications, and various measures of health care processes, such as implementation of effective treatment protocols.

Today, QI are being developed and used for a broad range of purposes in nursing homes and related facilities, such as assisted living. The federal government and many states plan to rely on QI as part of newer systems for monitoring and regulating nursing home quality of care. Although innovations that might improve nursing home quality are viewed positively, there are concerns about potentially inappropriate uses of QI. This paper describes the development of the federally sponsored nursing home QI, discusses a variety of applications of QI, and identifies issues and concerns regarding the use of QI.

PURPOSE

This paper is intended to help policymakers and others assess proposed uses of quality indicators; understand the current state-of-the-art, including issues such as reliability, validity, and risk adjustment; and understand the potentially promising uses and substantial limitations of quality indicators.

METHODOLOGY

Research for this paper consisted primarily of a detailed review of both published and unpublished literature and interviews with selected researchers and policymakers at the state and federal levels. Additionally, the author drew upon the records of meetings and conversations with participants who attended those meetings.

ORGANIZATION OF THE PAPER

The findings in this paper are presented in two key chapters and four appendices. The first chapter provides an overview of the development of the federally sponsored QI and a close look at the University of Wisconsin Center of Health Services Research and Analysis (CHSRA) QI that were developed in the early 1990s. That set of quality indicators is based on the Minimum Data Set (MDS), a federally mandated resident assessment tool designed to assess residents’ status when they enter a nursing home and regularly thereafter. This chapter describes how QI work and identifies key issues to consider when drawing conclusions from the CHSRA data.

The second chapter chronicles the development of a second set of quality measures, the “Mega QI” as they came to be called, and their use in a federally sponsored project to make facility-specific nursing home quality information available to the public. This chapter explores reliability, risk adjustment, and validity, important issues relating to the public disclosure of quality measures.

The four appendices detail specific applications of QI. Appendix A examines the federal government’s use of QI for targeting potential problems for further investigation in the federal nursing home survey process. Appendix B reviews the use of QI in internal quality
improvement. Appendix C details the use of QI in consumer information systems, and Appendix D describes efforts to link QI and public payment.

FINDINGS AND CONCLUSIONS

The current QI systems are the direct result of a 1986 Institute of Medicine (IoM) report, *Improving the Quality of Care in Nursing Homes*. Research generated by the report led to the development of the Minimum Data Set (MDS) resident assessment instrument, which in turn led to the development of a set of MDS-based quality indicators (the CHSRA QI) in conjunction with a four-state demonstration project. This demonstration project led to the selection of 24 QI for use in the federal nursing home survey process. These QI were designed to be a tool for capturing potential problems in performance that would highlight certain facilities for additional review.

The validity and reliability of these QI have been tested to some extent in a number of relatively small studies. Initial evidence on the reliability of the underlying MDS was encouraging, although more recent findings have raised some questions about the reliability of the MDS as it is used in actual practice. Limited research has also raised some questions about the reliability of several of the QI themselves. The results of the small number of studies on the validity of the QI have been mixed.

Three methodological issues are important when making inferences about quality from QI: risk adjustment, facility discharge practices as they relate to short-stay patients, and smaller facilities and rarer outcomes. Each of these issues can affect the results when QI are used to assess nursing home performance.

In April 2002, the Centers for Medicare and Medicaid Services (CMS) began publicly reporting a set of quality measures (QM) for nursing homes in six states on its website, “Nursing Home Compare.” The agency expanded this reporting to the national level in November 2002. CMS uses the term “quality measures” instead of “quality indicators” because the agency believes the new measures have been sufficiently validated to qualify as true measures of nursing home quality, as distinguished from the survey process QI, which are to be used as pointers to potential quality problems.

CMS contracted with Abt Associates (a Boston-based consulting firm). This large-scale, multi-million dollar research project is known as “the Mega QI project.” In addition, CMS contracted with the National Quality Forum (NQF) to make recommendations to CMS regarding measures to be used in CMS’ public reporting efforts. These two related projects have resulted in an explosion of research, analysis, and debate regarding QI in nursing facilities.

One prominent feature of the Abt research teamwork was the application of three types of risk adjustment to the QM:

- To reduce the chance that a nursing home that serves more frail residents is disadvantaged by the quality measures, information about some residents may be left out of the quality measures.
- All but one of the QM are statistically adjusted for resident clinical characteristics that predispose a patient to have a particular clinical problem (e.g., pressure sores).
- Three of the measures are presented with a special risk adjustment, the “facility admission profile” (FAP). This adjustment factor was applied to calculating a quality indicator at the facility level to address two issues that CMS’ contracted researchers believed would distort a facility’s quality rating: differential measurement practices (how well nursing home staff are able to complete MDS assessments) and differential selection
(the types of patients the home admits). The FAP has proved to be quite contentious, although CMS has decided use it in its new national reporting system.

The Mega QI research team conducted a validation study on the new QM and concluded that many of the quality indicators capture meaningful aspects of nursing facility performance and are reliably measured.

Methodological issues affecting the use of QI in assessing nursing home performance and the debate accompanying the Mega QI and NQF process only serve to underscore this paper’s major conclusion – that users must exercise caution when using and interpreting these complex measurement tools.

Designing, using, and evaluating QI involve a complex process of balancing tradeoffs or constraints. The optimal balance depends on the intended purpose and audience. For example, the information best suited for internal quality management or for the nursing home survey process is not necessarily best suited for purchasers such as Medicaid or for individuals choosing a nursing home.

The recent explosion of research has contributed substantially to the understanding of QI and their appropriate applications. However, it is clear that the development and use of quality indicators is still an evolving science. The challenge for policymakers and users is to apply the results of these studies wisely with an appropriate matching of the stated purpose against the strengths and limitations of QI.
INTRODUCTION AND BACKGROUND

1.0 INTRODUCTION

Quality indicators (QI) are markers of potentially poor or excellent health care quality. They may include such elements as mortality rates, avoidable complications, and various measures of health care processes, such as implementation of effective treatment protocols. The first generation of federally sponsored quality indicators in nursing homes was developed in the early 1990s by researchers at the University of Wisconsin Center for Health Services Research and Analysis (CHSRA). That set of quality indicators is based on the federally mandated Minimum Data Set (MDS), a comprehensive clinical and functional assessment instrument used in the federal nursing home survey process. The CHSRA quality indicators consist of a set of specific markers, including the prevalence of falls and the use of physical restraints on residents.

Today, QI are being developed and used for a broad range of purposes in nursing homes and related facilities, such as assisted living. The federal government and many states plan to rely heavily on QI as part of newer systems for monitoring and regulating nursing home quality of care. Although most applaud increased attention to innovations that might improve nursing home quality, concerns have been raised about potentially inappropriate uses of QI.

This paper is intended to help policy analysts and others:

- Assess proposed uses of quality indicators as part of:
  - public oversight systems,
  - private accreditation systems,
  - nursing home management and quality improvement systems,
  - consumer information systems such as report cards, and
  - payment systems intended to reward high performers;
- Understand the current state-of-the-art, including issues such as reliability, validity, and risk adjustment; and
- Understand the strengths and limitations of quality indicators.

1.1 METHODOLOGY

Research for this report consisted primarily of a detailed review of the literature and interviews with selected researchers and policymakers at the state and federal levels. In addition to books and articles, the literature search included a computer-assisted search of Medline and HealthStar, and a review of both published and unpublished government documents. Additionally the author drew upon the records of meetings and conversations with meeting participants.

1.2 QUALITY VERSUS QUALITY INDICATORS

Advocates and public officials have documented serious quality problems in facilities caring for older persons numerous times over the past seventy years. In 1927, federal surveyors were appalled by conditions in public almshouses, noting that the “inmates” frequently had inadequate food and clothing and that frail, elderly widows were inappropriately housed with “criminals, imbeciles, and the insane.”1 In 1998—three decades after federal standards for

---

1 B. Manard, R. Woehle, and J. Heilman, Better Homes for the Old (Lexington, Massachusetts: Lexington Books,
Medicare- and Medicaid-financed nursing homes were established with the passage of those financing programs—the U. S. General Accounting Office (GAO) investigated reports of serious problems in California nursing homes. GAO concluded, “Unacceptable care continues to be a problem in many homes,” noting that its investigators found homes with a pervasive smell of urine, residents with soiled bedding and clothing, inadequate diets, and untreated fractures. Each of those problems was pointed out as evidence of poor quality by concerned investigators. “Quality,” all agree, is a multi-dimensional concept. Defining it is one way to understand it. The dictionary defines quality as “excellence in kind.” Applying that definition to nursing facilities requires deciding what kind of places they are: What are nursing facilities supposed to do and be? A 1972 federally sponsored national study of nursing home quality noted the difficulty of the task, given the conflicting goals of these organizations even then. The study stated that, at a minimum, as organizations, nursing facilities had to be financially viable to survive. As places primarily charged with caring for frail older persons and substantially supported by public health care program funds, nursing homes were expected to “provide services that insure hygiene and biological functioning (health maintenance) and provide a nice place to live (quality of life).” Over time, as nursing homes have changed, the emphasis on the relative importance of one or the other aspect of the expected nursing home product has shifted; now the difference between quality of life and health is beginning to blur with greater appreciation of the interplay between the two. But still today, federal policy analysts note the difficulty of “defining the nursing facility product” sufficiently clearly to develop strategies for paying appropriately for nursing facility care and monitoring quality.

Measuring quality is a related but different way to put tangible bounds around this multi-dimensional concept. Measuring quality requires selecting and defining specific aspects of quality, developing ways to quantify (scale) differences, and sometimes determining standards (cut-points) to distinguish between acceptable and unacceptable quality. Quality indicators are one particular approach to measuring quality. However, designing, using, and evaluating proposed quality indicators involve balancing tradeoffs or constraints. The optimal balance depends on the intended purpose. The remainder of this section discusses some of the complexity involved in matching the intended purpose with constructing and using QI.

In 1998, the Health Care Financing Administration (then HCFA, now “the Centers for Medicare and Medicaid Services” [CMS]) sponsored a national conference of policymakers, purchasers, quality measurement researchers, consumer organizations, and other leaders in the

1977).
4 For example, if one thinks of nursing facilities principally as health care providers, then one might focus attention on creating ways to measure quality by starting with this definition: “Quality health care means doing the right thing, at the right time, in the right way, for the right person—and getting the best possible results” (U.S. Department of Health and Human Services, [DHHS], Your Guide to Choosing Quality Health Care (Agency for Health Care Policy and Research [AHCPR] Pub. No. 99-0012, December 1998). But if one also thinks of a nursing facility as a place where people live, then the applicable aspects of quality may be less related to what health care staff do to residents, and more related to opportunities for privacy and meaningful social engagement with the surrounding community.
5 B. Manard, R. Woehlke, and J. Heilman, Better Homes for the Old.
Participants agreed that unsuccessful efforts to collect and use health care quality information frequently stem from failure to follow an elementary rule: selecting appropriate quality indicators requires first clarifying the purpose and audience for the information. Too often, participants agreed, quality indicators are developed or selected simply from information most readily available rather than from information relevant to the selected QI. For example, the information best suited for internal quality management and improvement is not necessarily the same as that most useful for public accountability to purchasers such as Medicaid or individuals choosing a nursing facility.

Table 1.1, adapted from a useful chart prepared for the HCFA conference, summarizes some important differences between the two major uses for QI: quality improvement and public accountability (see Appendix B for more information on the issue of quality indicators for internal quality improvement initiatives).

---

<table>
<thead>
<tr>
<th>Issue</th>
<th>Measurement for Quality Improvement</th>
<th>Measurement for Public Accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Purpose</td>
<td>Identify process to be improved, or test results of efforts</td>
<td>Make a purchase decision, evaluate programs and public policies, provide reassurance, or conduct needs assessments</td>
</tr>
<tr>
<td>2. Requester/Audience</td>
<td>Internal (e.g., staff, providers, or managers)</td>
<td>External (e.g., purchasers, payers, accrediting agencies, regulators, policymakers)</td>
</tr>
<tr>
<td>3. What to measure</td>
<td>Biggest gap between practice and science</td>
<td>Measures with wide public acceptance or importance</td>
</tr>
<tr>
<td>4. Frequency of measure</td>
<td>Very frequent (feedback daily, weekly, or monthly)</td>
<td>Less frequently (e.g., annually)</td>
</tr>
<tr>
<td>5. Comparison</td>
<td>Longitudinal, within one facility or unit, or external for benchmarking</td>
<td>Cross-sectional (across units) or longitudinal</td>
</tr>
<tr>
<td>6. Sample size</td>
<td>Often relatively small</td>
<td>Large sample sizes with small confidence intervals</td>
</tr>
<tr>
<td>7. Unit of analysis</td>
<td>Smallest relevant unit (e.g., measuring the amount of lunches actually consumed in a particular facility when a different menu or approach to food choices is introduced)</td>
<td>Aggregate</td>
</tr>
<tr>
<td>8. Severity (risk) adjustment</td>
<td>Often not necessary if you believe that the processes are changing more than relevant characteristics of the patients.</td>
<td>Critical for fairness</td>
</tr>
<tr>
<td>9. Detection of bias</td>
<td>No audit, measurement done internally</td>
<td>External audit and/or external measurement required</td>
</tr>
<tr>
<td>10. Level of sophistication</td>
<td>Simple, not likely to be challenged</td>
<td>Appears simple, but rigorous and defensible</td>
</tr>
<tr>
<td>11. Level of detail</td>
<td>Very specific, miniscule</td>
<td>Summarized, global</td>
</tr>
<tr>
<td>12. Expected response</td>
<td>Behavior change</td>
<td>Decision-making</td>
</tr>
</tbody>
</table>

The chart illustrates key differences in the types of measurements and systems that are appropriate for different purposes. For example, risk adjustment may not be important when quality indicators are used as part of an internal quality improvement effort (in which the facility is tracking its own performance over time, rather than comparing its performance on QI to other facilities); if managers believe that the types of residents they are caring for are not changing dramatically over time. In contrast, large sample sizes and appropriate risk adjustment are considered “critical for fairness” when QI are used in public accountability systems, such as report cards, which are developed to help consumers choose a nursing home.

---

The impulse to use readily available information to construct QI for nursing homes stems from a laudable desire to minimize data collection burden and expense, as well as a sense of urgency with regard to trying to improve nursing facility care. Participants at the HCFA conference noted: “With limited resources and limited but growing capacity of health care organizations to be ‘measured,’ it is critical that sponsors of quality data collection efforts prioritize and select carefully the quality measures they require.” Conference participants also noted key risks in relying on the most readily available information, if that information were not actually best for the defined purpose: “How quality is defined and the specific indicators used to represent quality performance send powerful messages to health care providers and organizations regarding aspects of care to focus on improving. Selection of aspects of quality less pertinent to improving health could easily misdirect resources for quality improvement.”

Considerable confusion has arisen with regard to the use of quality indicators in nursing homes for purposes beyond those designated in the federal nursing home survey process, such as for quality improvement or public accountability (Table 1.1). The nursing home survey is intended to “help insure that nursing homes maintain compliance with federal standards” (see the following section for more detail on the nursing home survey process).9 Knowing that a nursing facility meets certain standards—alone or in particular combinations—may or may not be an optimal measure of quality for other purposes such as identifying important care processes that need improvement.10 Similarly, monitoring facility performance on a limited set of QI is not a quality assurance system. Two distinctions between a quality assurance system and other uses for quality information are important.

First, as CMS’s former Administrator has noted, “One can monitor quality at the level of aggregate profiles, but quality must be assured in every individual case…. The level of aggregation and statistical heterogeneity that may be quite acceptable for [quality monitoring and other purposes] is far too blunt and imprecise for quality assurance…. The problem of moving from the statistical norm to the random individual case constitutes the central dilemma of quality assurance.”11

Second, many standards in the federal nursing home quality assurance system are necessarily related to indirect care processes—proper food handing, for example. “Quality” is a complex concept; only aspects of quality can be measured. Ultimately, the primary concern is that nursing home residents are cared for well. Thus, some argue, the focus of a quality assurance system should be on outcomes.12 Did the residents attain or maintain the highest practicable level of health and functional status? Are the residents sufficiently satisfied with their quality of life? Resident care outcomes are important, but they are not always the best approach to standards. For example, few would argue against having standards for proper food handling or fire safety in a publicly certified health care facility. Nor would many argue that monitoring the incidence of food poisoning or fatal fires (outcomes) after the fact is an appropriate approach to quality assurance.13

---

10 See the following chapter for a discussion of using QI to improve quality.
13 For additional discussion of the benefits of process measures in both quality monitoring and assurance systems,
1.3 DEVELOPMENT OF FEDERALLY SPONSORED QI, BASED ON UNIFORM RESIDENT ASSESSMENTS

The perspectives embedded in today’s generation of nursing home quality indicators strongly reflect the issues that were most important when development first began as part of an effort to improve nursing home care two decades ago.

The federally mandated uniform resident assessment system and subsequent quality indicator systems built on MDS data are the direct result of a report, *Improving the Quality of Care in Nursing Homes*, issued in 1986 by the Institute of Medicine (IoM). The report was commissioned by HCFA in 1982. The IoM was charged with undertaking a study that would “serve as the basis for adjusting federal (and state) policies and regulations governing regulation of nursing homes so as to make those policies and regulations as appropriate and effective as possible.”

The IoM concluded that the three central requirements for providing high-quality nursing were: “(1) a competently conducted, comprehensive assessment of each resident; (2) development of a treatment plan that integrates the contributions of all the relevant nursing home staff, based on the assessment findings; and (3) properly coordinated, competent, and conscientious execution of all aspects of the treatment plan. The assessments should be repeated periodically and the treatment plan adjusted periodically.”

The IoM’s emphasis on care assessment and planning responded to complaints over the previous decade that nursing home standards and federal surveys needed to become more resident-oriented, and less facility-oriented.

The IoM proposed sweeping reforms, most of which became law with passage of The Nursing Home Reform Act, known as “OBRA ’87.” Regulations implementing the law (amended in 1988, 1989, and 1990) were promulgated over the following decade. The regulations:

- **Defined Quality Care**, stating that:
  “Each resident must receive and the facility must provide the necessary care and services to attain or maintain the highest practicable physical, mental, and psychosocial well-being, in accordance with the comprehensive assessment and plan of care”

- **Specified Important Elements of Quality**, including the following:
  “The resident has the right to be free from any physical or chemical restraints imposed for purposes of discipline or convenience, and not required to treat the resident’s medical symptoms.”

  “The facility must conduct initially and periodically a comprehensive, accurate, standardized, reproducible assessment of each resident’s functional capacity.”

---


16 The Omnibus Budget Reconciliation Act (OBRA) of 1987 amended the Social Security Act with respect to federal certification of nursing facilities.

17 Requirements of Participation (ROP), Section 483.25.

18 ROP Section 483.13(a).

19 ROP Section 483.20.
“The facility must develop a comprehensive care plan for each resident that includes measurable objectives and timetables to meet a resident’s medical, nursing, and mental and psychosocial needs that are to be identified in the comprehensive assessment.”\textsuperscript{20}

Research to develop the uniform assessment instrument and process began in 1988, led by Dr. Catherine Hawes.\textsuperscript{21} The assessment instrument (the Minimum Data Set or MDS) requires nursing facility staff to indicate (by checking the appropriate box on a form) a resident’s functional status (e.g., how much help the resident receives with toileting) and other conditions. The care-planning and assessment process includes Resident Assessment Protocols (RAPS) and “Triggers.” When a resident’s assessment reveals one or more of 18 indicators of potentially problematic conditions (e.g., incontinence), it triggers a required set of additional care-planning activities designed to address the problem.

The initial version of the assessment instrument was implemented in 1990 and has now been replaced by a second generation of assessment instrument and care-planning protocols that have been implemented in nursing facilities nationwide.\textsuperscript{22} In addition, electronic transmittal of MDS data to HCFA, initially mandated to begin in mid-1998, is now substantially operational.

In 1989, a team of researchers—led by Dr. David Zimmerman of The Center for Health Systems Research and Analysis (CHSRA) at The University of Wisconsin—began work on a set of quality indicators that would be based on the data gathered through the MDS. Much of the development was carried out in conjunction with a four-state demonstration project: The National Case-Mix Reimbursement and Quality Demonstration.\textsuperscript{23} The Wisconsin researchers

\textsuperscript{20} ROP Section 483.20(k).


\textsuperscript{22} In 1990, an early version of the assessment instrument and process was implemented by a transmittal from CMS (formerly HCFA) to the states. Proposed rules to formally establish requirements concerning the use of the MDS were published in December 1992. A revised version (MDS 2.0) was phased in across the nation and implemented in most homes by 1996.

\textsuperscript{23} In 1989, CMS (formerly HCFA) issued an invitation to all states to submit a proposal to participate in the demonstration project. Very few states submitted proposals. From those few, four states were selected: Maine, Kansas, Mississippi, and South Dakota. Texas, New York, and Nebraska (to a much lesser extent) also participated.
also created reports and software for facilities to use as internal performance improvement tools. The original purpose of these “CHSRA QI,” stemming from the IOM report, was to provide a tool for focusing surveys on facilities with the greatest probability of having serious problems. The next section describes the CHSRA QI in some detail, examining technical issues that are important to understand when considering any nursing home QI.

in aspects of the demonstration. The demonstration got off the ground in approximately 1991 and ended in 1998. The final evaluation was completed in 2000 but had not yet been released as of early November 2002. 24 IoM recommended that, “the standard survey would be designed to use ‘key indicators’ of performance to identify facilities with poor resident outcomes that might have resulted from substandard nursing home performance. If a facility had problems on the key indicators, it would be subjected to an extended survey protocol….” (IoM, *Improving the Quality of Care in Nursing Homes*).
2.0 INTRODUCTION

On July 1, 1999, a set of 24 federally sponsored quality indicators (QI) based on data from the Minimum Data Set (MDS) became part of the official nursing home survey process. Those 24 CHSRA QI are a subset of the original QI developed by researchers from the University of Wisconsin Center for Health Services Research and Analysis. This chapter describes the 24-item CHSRA QI and discusses some key issues in interpreting them.

2.1 OVERVIEW OF THE 24-ITEM CHSRA QI

Table 2.1 lists the 24 CHSRA QI used in the new federal nursing home survey process. An algorithm based on residents’ MDS quarterly assessments defines each CHSRA QI. Most (20) of these QI are prevalence measures. They give a point-in-time snapshot of the percentage of residents in a facility with a particular condition. For example, one quality indicator—“prevalence of indwelling catheters”—is defined as the number of residents with an indwelling catheter divided by the total number of residents. The remaining four QI are incidence measures. They are measures of the number of new occurrences of particular conditions that developed from one assessment period to the next. For example, the “incidence of cognitive impairment” is defined as the number of residents who were newly “identified as being” cognitively impaired on the most recent assessment divided by the number of residents who were not cognitively impaired on the previous assessment.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Quality Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents</td>
<td>1. Incidence of new fractures</td>
</tr>
<tr>
<td></td>
<td>2. Prevalence of falls</td>
</tr>
<tr>
<td>Behavioral/Emotional Patterns</td>
<td>3. Prevalence of behavioral symptoms affecting others (verbally abusive, physically abusive, or socially inappropriate/disruptive behavior) [Risk Adjusted]</td>
</tr>
<tr>
<td></td>
<td>4. Prevalence of symptoms of depression (sad mood plus at least 2 of the following: resident made negative statements, agitation or withdrawal, wakes with unpleasant mood, suicidal or has recurrent thoughts of death, weight loss)</td>
</tr>
<tr>
<td></td>
<td>5. Prevalence of symptoms of depression and no antidepressant therapy</td>
</tr>
<tr>
<td>Clinical management</td>
<td>6. Prevalence of residents using 9 or more different medications</td>
</tr>
<tr>
<td>Cognitive patterns</td>
<td>7. Incidence of cognitive impairment</td>
</tr>
<tr>
<td>Elimination/incontinence</td>
<td>8. Prevalence of bladder or bowel incontinence [Risk Adjusted]</td>
</tr>
<tr>
<td></td>
<td>9. Prevalence of occasional bladder or bowel incontinence without a toileting plan</td>
</tr>
<tr>
<td></td>
<td>10. Prevalence of indwelling catheters</td>
</tr>
<tr>
<td></td>
<td>11. Prevalence of fecal impaction</td>
</tr>
<tr>
<td>Infection control</td>
<td>12. Prevalence of urinary tract infections</td>
</tr>
<tr>
<td>Nutrition/eating</td>
<td>13. Prevalence of weight loss</td>
</tr>
<tr>
<td></td>
<td>14. Prevalence of tube feeding</td>
</tr>
<tr>
<td></td>
<td>15. Prevalence of dehydration</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>16. Prevalence of bedfast residents</td>
</tr>
<tr>
<td></td>
<td>17. Incidence of decline in late loss ADLs</td>
</tr>
<tr>
<td></td>
<td>18. Incidence of decline in range of motion</td>
</tr>
<tr>
<td>Psychotropic drug use</td>
<td>19. Prevalence of antipsychotic use in the absence of psychotic and related conditions [Risk Adjusted]</td>
</tr>
<tr>
<td></td>
<td>20. Prevalence of antianxiety/hypnotic use</td>
</tr>
<tr>
<td></td>
<td>21. Prevalence of hypnotic use more than two times in the last week</td>
</tr>
<tr>
<td>Quality of life</td>
<td>22. Prevalence of daily physical restraints</td>
</tr>
<tr>
<td></td>
<td>23. Prevalence of little or no activity</td>
</tr>
<tr>
<td>Skin care</td>
<td>24. Prevalence of stage 1-4 pressure ulcers [Risk Adjusted]</td>
</tr>
</tbody>
</table>

Four of the 24 CHSRA QI have associated risk adjusters. A risk adjuster ideally adjusts for residents’ health and functional conditions—beyond a facility’s control—that place a resident at greater risk for incurring a condition that triggers a quality indicator. For example, some residents enter a nursing facility so sick and debilitated that they are far more likely to develop a pressure ulcer than are other residents, even in a facility with excellent care. Therefore, to compare potential quality problems with respect to pressure ulcer prevention at one facility to another (or at one facility over time), it is important to understand how the homes differ with respect to the characteristics of their residents (the relevant “case-mix”). In the CHSRA QI with risk adjusters, residents are divided into two groups: high risk and low risk. For example, for the
risk adjuster associated with QI #24—“prevalence of stage 1-4 pressure ulcers”—the two groups are:

- **High Risk:** Residents with impaired transfer or bed mobility or comatose or malnutrition or end stage disease.
- **Low Risk:** All others

Risk adjusters used in the CHSRA QI are, of course, limited to conditions that can be determined from the MDS. Further, the 24-item CHSRA QI used in the survey process are limited to information found on the quarterly assessments. Some would argue that an additional complication is that almost all conditions that residents have after they have been in a nursing home for a while could be the result of the care process itself and thus, only conditions identified on admission should be used as part of the definition of risk adjusters used with quality indicators. For example, “malnutrition” is used as one of the factors in the CHSRA QI risk adjuster related to pressure ulcers, but malnutrition in a long-stay resident likely signifies something different than malnutrition in a newly admitted resident.

2.2 ASSESSING HOW WELL THE CHSRA QI MEASURE QUALITY OR ITS COMPONENTS

This section discusses some key issues regarding how well the CHSRA QI measure the aspects of quality that are included in the set of 24 items discussed above. The researchers who developed the CHSRA QI have repeatedly stressed that “Quality indicators are not direct measures of quality; they are pointers that indicate potential problem areas that need further review and investigation.” The researchers clearly warn regulators and providers against the misuse of the CHSRA QI, cautioning that, “State surveyors will be able to access QI reports…[but] they will also need to carefully investigate the potential problem areas…to determine whether deficiencies should be issued.” They also caution providers to “be careful to not use the information in your QI reports to draw specific comparisons to other facilities as a final determination of quality requires detailed clinical review and investigation.”

But danger is always present that those warnings will be lost. The widespread availability of officially sanctioned, computer-generated numbers labeled “Quality Indicators” can in itself create an impression of certainty and precision. Thus, it is critically important to look beneath the numbers and ask what they mean. For example, if one nursing home scores half as well on a particular quality indicator as another, does that mean that one home is half as good as the other with respect to quality in general or even to the specific aspect of quality addressed by the quality indicator used?

The quality of measurement tools in general (like the MDS and the CHSRA QI) is judged by well-established standards based on tests of various characteristics ("psychometric properties"), notably reliability and validity. In addition, special considerations, such as

---

26 The specific caveats from Center for Health Systems Research and Analysis (CHSRA) cited in this section were on the organization’s web site (http://www.chrsa.wisc.edu/CHSRA/QIs/qivsmeas.htm) from at least May 1998, through at least April 2000. The web site has subsequently changed, although CHSRA continues to caution that the QI are pointers that indicate potential problems.
27 Ibid.
28 Ibid.
appropriate risk adjustment, are relevant to the quality of measurement tools in various applications. Psychometric properties as well as risk adjustment help determine how much confidence should be placed in inferences drawn from a nursing home’s performance on the CHSRA QI.

One analysis found that only about half of the general health care quality indicators reviewed had been subjected to appropriate reliability and validity studies. The CHSRA QI and the MDS on which they are based have been tested to some extent and HCFA has stated its “commitment to change” over time as problems are revealed. Notably, however, remarkably few studies of the psychometric properties of the MDS or CCHSRA QI have been conducted. Most of these studies were relatively small. Further, most were conducted as part of the development work, using specially trained assessors. This contrasts sharply with the extensive research available on other types of assessment tools. Findings from the limited studies of MDS and the CHSRA QI are discussed below.

2.2.1 Reliability

“Reliability” refers to the consistency of measurements produced when an instrument is used. Reliability is defined as “the degree to which an instrument measures the same way each time it is used under the same conditions with the same subjects.” For example, most inexpensive weight scales used in homes are less reliable than the scales that physicians use. If one steps on an inexpensive home scale five times, the readings may vary, perhaps by several pounds. More closely similar readings from an inexpensive weight scale are more likely if one is very careful to step on the scale in exactly the same spot each time. But, in general, the better-made scales used in physicians’ offices register the same weight each time with less sensitivity to minor variations in how a person stands on the scale. Similarly, a well-designed assessment instrument will produce more reliable results consistently than one that is less well designed. Ambiguously worded questions are a common source of unreliability in an instrument. In addition, questions that require subjective judgment or information that is difficult to obtain tend to provide less reliable answers.

Researchers typically assess several different aspects of an instrument’s reliability. One of the most important for conclusions drawn from the CHSRA is inter-rater reliability—“the degree to which two or more judges (raters) rate the same variables in the same way.” To assess the inter-rater reliability of an assessment instrument like the MDS or the CHSRA QI, researchers have two different people complete assessments on the same nursing home residents and then compare the results.

Reliability of the Underlying MDS Data

Because the CHSRA QI rely on MDS data, it is important that the MDS be reliable. Just four studies of inter-rater reliability for the MDS have been published in peer-reviewed scientific journals; all of these have very limited samples. The first reports on research conducted in the course of developing the original instrument; dual assessments were conducted on 123 residents


31 Ibid.
from 13 facilities. A second small study was conducted as part of refining the instrument to produce a later version; dual assessments were conducted on 187 residents. A third study reanalyzed data from the first two, focusing on the inter-rater reliability of assessments for cognitively impaired residents. In each of those three studies, nurses who had been specially trained by research staff conducted the assessments. Only one small study, the fourth of the inter-rater reliability studies, tested the instrument as it is actually used in practice by staff nurses without special training.

The results of these published studies, in brief, are as follows:

- In general, nearly all items on the MDS met or exceeded acceptable standards for inter-rater reliability in the published studies. At least one item used in the CHSRA, however, failed to meet acceptable standards: presence of “end stage disease.”
- Inter-rater reliability was lower—though still within acceptable levels—for the general nursing home population in the one small (33 residents in one nursing home) study that used regular staff nurses to do the assessments.
- Inter-rater reliability on assessments for cognitively impaired residents was significantly lower than on other assessments.

In sum, the initial evidence was encouraging with regard to inter-rater reliability, except for assessments of those who are cognitively impaired, a growing proportion of the nursing home population. More recently, however, findings have emerged from an additional study that raises new questions about inter-rater reliability as the MDS is used in actual practice (i.e., by facility nurses, rather than by specially trained researchers).

If the underlying data on which CHSRA QI are based are unreliable, as some studies suggest, this increases the need to view results based on the CHSRA QI cautiously.

Reliability of the CHSRA QI Per Se

Some indication of the inter-rater reliability of the CHSRA QI is provided by the studies of the inter-rater reliability of MDS items. However, most of the CHSRA QI are defined by combinations of several individual MDS items. If four items in the definition of a QI each have good inter-rater reliability scores, the overall inter-rater reliability of the entire QI element still may not be adequate. No studies of the inter-rater reliability of the CHSRA QI have been studied as part of the evaluation of the nursing home case mix and quality demonstration and had not been released by CMS as of mid-October 2002 for more widespread consideration. There is an apparent disagreement among various parties regarding interpretation of study results. MedPAC, however, says the study points out notable errors in the MDS assessment forms completed by facility nurses, compared to assessments done by specially trained research staff.

One analyst offered the following example when discussing quality indicators in general. “Reliability engineers often calculate system failure probabilities in their work. These calculations are dependent upon a system’s architecture and design. If three serially connected components each had a 0.85 probability of failing in a year, their

---

36 [www.medpac.gov](http://www.medpac.gov), transcript of meeting, January 11, 2001. One study discussed by MedPAC staff was conducted as part of the evaluation of the nursing home case mix and quality demonstration and had not been released by CMS as of mid-October 2002 for more widespread consideration. There is an apparent disagreement among various parties regarding interpretation of study results. MedPAC, however, says the study points out notable errors in the MDS assessment forms completed by facility nurses, compared to assessments done by specially trained research staff.
37 One analyst offered the following example when discussing quality indicators in general. “Reliability engineers often calculate system failure probabilities in their work. These calculations are dependent upon a system’s architecture and design. If three serially connected components each had a 0.85 probability of failing in a year, their
published. However, as part of a large-scale project to develop quality measures for public reporting and other purposes (see the Mega QI project described in Chapter 3, “Additional Federally Sponsored QI and Other Recent Developments”), researchers have explored the reliability of a small set of CHSRA QI. The results have been made publicly available, but have not yet been formally peer-reviewed and published. This research team found that one of the nine CHSRA QI studied (“prevalence of little or no activity”) did not meet generally acceptable standards of reliability. In addition, the team’s clinical panel rejected two additional CHSRA QI (fecal impaction and dehydration) before conducting formal validation studies, in part because the panel felt that the reliability of each was compromised because at least one underlying MDS item was insufficiently robust.

2.2.2 Validity

“Validity” refers to “the degree to which an instrument measures that which it is supposed to measure.” For example, a weight scale obviously does not produce a valid measure of body fat percentage or cognitive function. But does QI #19 (“Prevalence of antipsychotic use in the absence of psychotic and related conditions”) produce a valid measure of the inappropriate use of antipsychotics? Many concepts in health care—such as “psychotic and related conditions”—are both complex and subject to dispute among specialists. Thus, validity is inherently difficult to assess. Researchers use various approaches to assessing validity; the more difficult the concept, the greater the need for sophisticated, well-designed studies in order to have confidence in the measurement tool. Confidence in the validity of the MDS and especially the CHSRA QI hinges almost entirely on “face validity” because relatively little relevant research on these federally mandated measurement tools has been conducted. Further, the available validity studies have revealed mixed results.

Face Validity and Content Validity

As noted above, the 24 CHSRA QI are a subset of QI initially selected by experts looking for reasonable indicators of quality that might be measured with the information available on the MDS. The items they chose do appear “on the face of it” to be reasonably related to aspects of nursing home quality and are thus said to have “face validity.” On the other hand, the 24 CHSRA QI do not address certain aspects of quality that experts and consumers consider important—staff attitudes and many elements of resident quality of life, for example. Thus, the CHSRA QI are limited with respect to “content” validity, that is, whether they measure all the aspects or “content” of quality.” The limited scope of the 24-item CHSRA QI is particularly

combined probability of failure would be 0.85 X 0.85 X 0.085 = 0.614” (E. Huff, “Comprehensive Reliability Assessment and Comparison of Quality Indicators and Their Components,” Healthcare Quality Analysis 50(12) (1998): 1395-1404.

38N. Sproull, 1995.

39Research is underway to address limitations of the CHSRA QI with regard to both short-stay Medicare patients (discussed in a subsequent section) and quality of life. CMS (formerly HCFA) has sponsored a project to develop quality of life measures and indicators as well as a catalogue of approaches that facilities can use to improve quality of life for nursing home residents. The project, led by Dr. Rosalie Kane of the University of Minnesota, is currently slated for completion in approximately 2003 (C. Graunke, “HCFA Nursing Home Update,” presented at the American Association of Homes and Services for the Aging, Spring Conference, Washington D.C., March 12, 2001). The study will develop QI based on data not available in the MDS; for example, researchers are focusing on environmental features of nursing homes such as private rooms. But the researchers will also test the quality of life measures against various MDS-derived indicators (R. Kane, personal communication, July 2000). For a summary discussion of quality of life issues and measures, see: R. Kane, “Review of Selected Research on Nursing Home
important in light of researchers’ findings that “quality” as measured on one indicator is rarely correlated with “quality” as measured on another.40

Testing Against a “Gold Standard”

“Criterion validity” is one of the key psychometric properties of instruments subject to formal testing. Two types of tests are used: testing a new instrument against an established one (a “gold standard”) and testing how well elements in a new instrument predict an outcome that logically should follow if the tested elements are valid.41 In 1995, researchers developing the CHSRA QI published a description of their development efforts.42 They noted, “Given the potential impact and intended uses of the QI, it is essential that their validity be established.” In that paper, they describe the validation studies that were underway. Researchers were to assess four to six QI in each of 20 homes. The study design involved comparing potential quality problems identified by the QI based on the homes’ MDS assessments to findings regarding “actual” quality problems as determined by experienced surveyors’ detailed, on-site review at each home. As of this writing, however, the results of those validation studies have not been published.43

Although it is somewhat surprising that the CHSRA QI system was used nationally to inform and improve the nursing home inspection process in the absence of large-scale published validity studies, a handful of independent published studies relevant to the validity of eight CHSRA QI had been conducted before 2001. These studies either compared the CHSRA QI to other established scales or for one QI, incontinence, measured how well the QI predicted the actual outcome. As can be seen in Table 2.2 the results of those studies are mixed.


40 For example, one large-scale study of three quality indicators in nursing homes found that “once risk adjustment was taken into account, [these] quality measures that were based on different outcomes were not correlated, therefore one cannot assume that the relative rank of a nursing home is the same for all outcomes. This supports conclusions reached by [other researchers] that one cannot generalize findings about provider quality from one disease to another” (D.B. Mukamel and C.A. Brower, “The Influence of Risk Adjustment Methods on Conclusions About Quality of Care in Nursing Homes Based on Outcome Measures,” Gerontologist 38(6) (1998): 695-703.

41 These two types of tests for “criterion validity” are referred to as “convergent validity” and “predictive validity.”


43 A 1998 article contained this short summary: “Findings from the validation tests showed that the quality indicators have a high degree of accuracy (reliability) and are useful tools for identifying quality problems related to all aspects of the nursing care process (i.e., assessment, care planning, implementation, and monitoring of care)” (S.L. Karon and D.R. Zimmerman, “Nursing Home Quality Indicators and Quality Improvement Initiatives,” Topics in Health Information Management 18(4) (1998): 46-58. Although the researchers completed their final report on the CHSRA QI validation studies, the report had not been made available to the general public through the federal repository of sponsored research (i.e., the National Technical Information Service) as of March 2001.
Table 2.2
Summary Results of Validity Studies Related to CHSRA QI Version 6.3 and Published Before 2001

<table>
<thead>
<tr>
<th>Quality Indicator</th>
<th>Results of Relevant Studies⁴⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Prevalence of behavioral symptoms affecting others</td>
<td>MDS behavior scale did not correlate well with established scales used in comparisons.⁴⁵</td>
</tr>
<tr>
<td>4. Prevalence of symptoms of depression</td>
<td>MDS depression scores were poorly correlated with established scales used in comparisons.⁴⁶</td>
</tr>
<tr>
<td>5. Prevalence of symptoms of depression without antidepressant therapy</td>
<td>MDS depression scores were poorly correlated with established scales used in comparisons (see above); Validity studies are not available for use of antidepressant therapy</td>
</tr>
<tr>
<td>7. Incidence of cognitive impairment</td>
<td>Various MDS measures of cognitive impairment were strongly correlated with comparison scales in each of 5 separate studies.⁴⁷</td>
</tr>
<tr>
<td>8. Prevalence of bladder or bowel incontinence</td>
<td>MDS ratings of incontinence were compared to wet checks performed by research staff. The 2 measures (incontinence ratings and wet checks) were poorly correlated in 2 of 9 homes but significantly correlated in the others.⁴⁸</td>
</tr>
<tr>
<td>9. Prevalence of occasional or frequent bowel or bladder incontinence without a toileting plan</td>
<td>The one study of incontinence validity had mixed results (see above). One study found that in an on-site investigation of a random sample of Texas nursing home residents, most (81%) of incontinent residents would have “potentially benefited from a toileting plan,”⁴⁹ which contributes to understanding this as a helpful QI, particularly if validity of the measure of incontinence itself were improved.</td>
</tr>
<tr>
<td>17. Incidence of decline in late loss activities of daily living (ADLs)</td>
<td>MDS measures of ADL function correlate well with established comparison scales in both of two separate studies; however, “reliability was suboptimal” in one study.⁵⁰</td>
</tr>
<tr>
<td>24. Prevalence of Pressure Sores Level 1-4</td>
<td>Although this is based on an established measure for grading pressure sores as they worsen, healing pressure sores do not correspond clinically to this scale. Several validated scales for healing pressure sores exist.</td>
</tr>
</tbody>
</table>

“Cognitive impairment” is the most studied element of those listed above and fared very well in five separate tests. None of the published studies defined “cognitive impairment” exactly

---

⁴⁴ Scales used in research studies cited do not exactly correspond to CHSRA QI definitions.
⁴⁹ L. Cortes, et al., A Statewide Assessment of Quality of Care, Quality of Life and Consumer Satisfaction in Texas Medicaid Nursing Facilities; Texas Department of Long Term Care, Office of Programs, Medical Quality Assurance, December 2000.
as it is defined in the CHSRA QI, but all of those studies did include at least one of the two MDS items used in the CHSRA QI definition. For example, in one study, researchers compared MDS ratings of one item related to cognitive status to assessments of the same 147 residents on two well-established instruments used to measure cognitive status—the Mini Mental Status Examination and the Test for Severe Impairment. The MDS item selected for this validity test was “resident’s ability to make consistent and reasonable decisions in everyday situations.” The researchers concluded that, “[The] results clearly demonstrate [a] high level of …validity of classifications derived from the MDS decision-making item with residents’ ratings on other well validated indicators of cognitive function.” MDS measures of activity of daily living (ADL) function also correlated well with established comparison scales in each of two separate studies. Validity studies of scales related to behavior, depression, and incontinence revealed less encouraging results, as indicated in Table 2.2.

### 2.2.3 New Research on the CHSRA QI

Two large-scale projects evaluating some of the CHSRA QI are contributing additional information, but full results of only one are publicly available at this writing. Even so, one can see that different study teams and different approaches can lead to different results. In one study, funded by the California Health Care Foundation, researchers evaluated the reliability and validity of a set of CHSRA QI. They found that three of the CHSRA QI examined (weight loss, being in bed all or most of the time, and use of restraints) were sufficiently reliable and valid to be used in California’s new web-based public reporting system as indicators of the comparative quality of facilities. Scientific papers have not yet been published but a preliminary description of the research is available.

In a second study, known as the “Mega QI study” and described in the next chapter, researchers determined that different criteria should be used to evaluate QI for different purposes (e.g., internal facility quality improvement versus public reporting); emphasizing particularly the need for the highest degree of validity for QI that were to be publicly available and intended to be seen as distinguishing among better or worse nursing facilities. In the first round of analyses, the researchers concluded that all 24 CHSRA QI were appropriate for internal quality improvement efforts, but that none of them, as defined by the original authors, was optimally suited for use in the survey process or for publicly reported measures, principally because better risk adjustment methods were required.

---

51 The CHSRA QI (Version 6.3) defines 2 levels of cognitive impairment based on how a resident is rated on 2 MDS items. “Cognitive Impairment” = “Impaired decision making” (MDS item B4 > 0) AND “short term memory problems” (MDS item B2a = 1). “Severe cognitive impairment” = “severely impaired decision making” (MDS item B4=3) AND “short term memory problems” (MDS item B2a=1).


53 This MDS item requires assessors to place residents at one of four levels: (1) Independent—resident’s decisions were consistent and reasonable; resident organized daily routine and made decisions in a consistent, reasonable, and organized fashion; (2) modified independent—resident organized daily routine and made safe decisions in familiar situations but experienced some difficulty when faced with new tasks or situations; (3) moderately impaired—resident’s decisions were poor; resident required reminders, cues, and supervision in planning, organizing, and correcting daily routines; and (4) severely impaired—resident’s decision-making was severely impaired; resident never (or rarely) made decisions.

54 http://www.calnhs.org/research
For example, the panel rejected the CHSRA QI related to mood (“symptoms of depression”) because they believed “the operational definition...identifies too few residents with depression so that reported rates are low and not indicative of the actually prevailing rates at a facility...only very astute assessors will document the necessary detail correctly.” The QI “prevalence of bedfast residents” was rejected in part because the panel concluded “there did not seem to be a strong rationale for a QI in this area” and because the measure appeared imprecise (i.e., “bedfast” does not represent residents who are totally dependent because residents could be lifted, placed in a recliner, or wheeled into a hallway). Of the other QI, the researchers reported that one (“prevalence of behavioral symptoms affecting others”) was not valid. Two others (“prevalence of anti-psychotic use in the absence of psychotic and related conditions” and “prevalence of physical restraints”) were reported to be insufficiently valid. One (“little or no activity”) was reported to be unreliable. Only five (“prevalence of bowel and bladder incontinence,” “prevalence of indwelling catheters,” “prevalence of urinary tract infections,” “incidence of decline in late loss ADLs,” and “prevalence of stage 1-4 pressure sores”) were reported to have a high degree of both reliability and validity.

2.2.4 Additional Issues in Drawing Inferences about Comparative Quality from QI

This section discusses three additional methodological issues that are important in making inferences about quality from QI.

**Appropriate Risk Adjustment**

As noted above, only four of the 24 CHSRA QI used in the new survey process have risk adjustments. Each of those four differs from the risk adjustments originally designed by the researchers; further, three additional risk adjustments originally designed for those 24 CHSRA QI could not be used in the survey process. These problems occurred because the quarterly assessments that policymakers decided to use for the mandated version (i.e., version 6.3) lack information that was available on the assessment forms used during development of the CHSRA QI.

Empirical studies of the 24 CHSRA QI and the risk adjustments actually used in the survey process have not been published. However, an early paper by CHSRA researchers illustrates both the importance of appropriate risk adjustment and the difficulty of that task. The CHSRA researchers used a combination of expert opinion and empirical analysis to develop the risk adjusters for the original CHSRA QI. Appropriate risk adjustment, as the researchers note, includes using judgment to determine when it is best not to use risk adjustment. For example, the researchers considered fecal impaction and dehydration to be such severe problems and so closely related to quality of care that “the possibility of overlooking care problems outweighs the potential benefits of risk adjustment.” Similarly, they say, it is important not to use items for risk adjustment that are themselves likely quality indicators within a facility’s control (e.g., use of restraints), even if these items are statistically related to an outcome of interest (e.g., falls and injuries).

55 K. Berg, et al., Identification and Evaluation of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings, p. 92.
56 Ibid., p. 58.
58 Ibid.
As noted earlier, the CHSRA QI researchers defined high and low risk groups for selected indicators. Facilities with CHSRA QI rates at or above the 90th percentile were defined as “outliers” with respect to potentially poor quality. The researchers found that “On average…42 percent of outlier facilities shifted from non-outlier status as a result of risk adjustment.” That is, without risk adjusting the QI, a very substantial proportion of the facilities in the study would have been misclassified as having potentially poor quality. The researchers also noted that the greatest shifts in facility outlier status occurred within high-risk groups for incontinence, the use of antipsychotic medications, and ADL decline. The first two but not the last are risk adjusted in the 24 CHSRA QI used in the survey process. This suggests that particular caution might be warranted when drawing conclusions about whether a facility is potentially good or poor based on CHSRA QI #17, “Incidence of decline in late loss ADLs.”

Research by various investigators further emphasizes the importance of risk adjustment to drawing appropriate conclusions about facility performance from quality indicator data. Numerous studies have found that hospital quality rankings on various indicators vary according to whether the QI is risk adjusted and, if so, how the risk adjuster is constructed and applied. Similar findings on nursing home quality indicators are emerging.

One set of researchers studied 550 nursing homes in New York State, using data from the State’s assessment instrument, which is similar to the MDS. They compared findings about those nursing homes with regard to three quality indicators: decline in functional status, worsening pressure ulcers, and prevalence of restraints. For each comparison among nursing homes, the researchers tested results using (1) no risk adjustment, (2) a simple risk adjuster, and (3) a sophisticated risk adjuster. Using the sophisticated risk adjuster as the standard, the researchers found that a large percentage of homes, as many as half in some tests, were misclassified (as “good” or “poor” quality homes) when the data were not appropriately risk adjusted. The researchers concluded that “Insufficient risk adjustment of outcome measures may…lead to inappropriate classification of nursing homes as either poor quality or high quality homes.” Researchers who studied rankings based on pressure ulcer development using data from over 30,000 residents in Department of Veterans Affairs (VA) facilities found similar results.

Facility Discharge Practices and Short-stay Residents

Using quality indicators to compare facility performance also requires attention to a factor researchers call “attrition bias.” For example, consider the use of CHSRA QI #24, “prevalence of stage 1-4 pressure ulcers.” Researchers have shown that residents who develop pressure ulcers are at substantially greater risk of general deterioration and earlier death than are others, in part because poor health itself is a risk factor for developing pressure ulcers. If one

---

59 The researchers chose to use high and low risk groups, rather than another approach to risk adjustment, in part because the chosen method “is a relatively simple and straightforward process that those who do not have extensive training in epidemiology or multivariate statistics can understand” (Ibid).

60 Ibid.

61 Additional concerns about this particular QI arise from apparent reliability problems with the MDS ADL scale when used by nursing home staff in the normal course of business, rather than by specially trained researchers.


nursing home tends to send worsening residents to the hospital far more quickly than does a second home, then that second home is more likely to have a higher “prevalence of stage 1-4 pressure ulcers” than the former, even if the second home actually does a better job with skin care. Sophisticated risk adjustment of quality indicators can mitigate, but not eliminate this problem, as was shown in a large study of nursing home residents treated in the VA system.65

Using the CHSRA QI to compare the performance of homes that specialize in short-stay residents to other homes raises similar issues. Some problems potentially resulting from poor quality nursing home care may not show up until after a resident has been discharged, and many problematic cases may not be caught on the quarterly assessments used in the 24-item CHSRA QI. Thus, some researchers recommend that optimally one should use episodes of care with a fixed exposure time when comparing health care outcomes.66 For example, federal hospital mortality statistics account for all deaths within 30, 90, and 180 days of admission, regardless of when the discharge occurred.

**Smaller Facilities and Rarer Outcomes**

Each of the conditions measured in the CHSRA QI can be thought of conceptually as the result of three factors: resident risk factors beyond a facility’s control (e.g., age or condition upon admission), quality of care in the nursing home, and random chance. Thus, even if risk factors were perfectly accounted for, homes with identical quality of care may perform quite differently on quality indicators due to random variations. This “background noise” is a bigger problem in making inferences from CHSRA QI with small facilities and rare events.67 For statistical analyses, pressure ulcers are a relatively rare event, even though they are found in about 10 percent of nursing home residents.68 Some researchers have concluded that to avoid excessive background noise, “limiting analyses to facilities with more than 100 patients may be the best approach.”69 Notably, however, fewer than half of all nursing homes are that large.

---

67 Technically, the residents observed at any one time are a sample of all residents the home might treat. Thus, sample size issues affect the probability of false negatives and false positives due to chance.
3.0 INTRODUCTION

CMS provides facility-specific information on nursing home quality issues and some resident characteristics on its “Nursing Home Compare” web site (see Appendix C for more on the use of quality indicators in consumer information systems). In April 2002, CMS began publicly reporting a set of “Quality Measures” for nursing homes in six states70 on the “Nursing Home Compare” web site.71 This pilot project will be expanded to a national public reporting system, scheduled for launch in November 2002. CMS uses the term “quality measures” (QM) rather than “quality indicators” (QI) because CMS believes that the new measures have been sufficiently validated—a matter of considerable controversy—to qualify as true measures of nursing home quality, as distinguished from the QI used in the survey process, which are to be used as pointers to potential quality problems.72 Several of these QM are closely related, but not identical, to QI originally developed by CHSRA; some are based on QI originally developed by other researchers.

To develop these measures for public reporting and other purposes, CMS contracted with Abt Associates (a Boston-based consulting firm) which assembled a team of researchers from several organizations. This large-scale, multi-million dollar research project is known as “the Mega QI project.”

In addition, CMS contracted with the National Quality Forum (NQF) to make recommendations to CMS regarding measures to be used in CMS’ public reporting efforts. NQF is a non-profit membership organization “created to develop and implement a national strategy for health care quality measurement and reporting,” using a consensus process.73

These two related projects have resulted in an explosion of research, analysis, and debate regarding QI in nursing facilities. Both the Mega QI project and the NQF process were ongoing as of this writing, although CMS has announced the measures to be used for the national public reporting initiative to begin in November 2002. This chapter outlines the process by which those measures were developed and highlights key issues in the debate.

3.1 DEVELOPMENT PROCESS

Shortly after Thomas Scully was confirmed as the current Administrator of CMS, he announced a commitment to make quality rankings of nursing homes available to the public with a goal of stimulating improvements in quality through market forces. Work of the Mega QI team (which began before the new administration took office) and the National Quality Forum (on whose Board Mr. Scully serves) were driven in substantial part throughout 2001 and early 2002 by CMS’ push to fulfill that commitment quickly.

The National Quality Forum convened a “Nursing Homes Performance Measures Steering Committee” in October 2001 to begin deliberations.74 At that point, the Mega QI

70 Colorado, Florida, Maryland, Ohio, Rhode Island, and Washington.
72 Provider organizations have proposed that CMS’ new QM should be called “quality markers,” rather than “quality measures.” In its reports, CMS’ research contractors most frequently use the term “quality indicators,” but have sometimes referred to these as “quality indicators (performance measures).” Debates about terminology reflect underlying differences and concerns regarding what quality indicators actually represent.
74 The eleven-member Steering Committee includes 2 prominent researchers in the field: Drs. R. Kane and D.
researchers had completed the first phase of their work: a detailed analysis of all existing MDS-based QI, development of a new risk adjustment method, and a preliminary test of those QI recommended by their clinical panel for further study. Plans for an expanded validation study (ultimately completed in August 2002) were underway.

Facing a deadline to make recommendations to CMS on measures to be used for the six-state pilot project, the NQF held numerous telephone conferences and advanced the date of their first public meeting to November 12, 2001. Two days later, NQF forwarded its steering Committee’s recommendations for seven chronic care and four post acute care measures for the pilot project. NQF cautioned CMS that, although the Steering Committee had limited its recommendations to MDS-based QI studied by the Mega QI team, “no such limitation will exist in its consideration of measures to recommend to the NQF membership….The Steering Committee agreed that other high priority issues regarding nursing home quality information exist but are not addressed by the ‘Abt indicators,’ [including] staffing information…, nursing home environment (e.g., sharing rooms, private baths, food quality), and patient autonomy.”

In addition, NQF cautioned CMS that, “Several members of the Steering Committee remain very concerned about Abt’s risk-adjustment methodology. Given CMS’ time constraints for measures to be used [in the pilot project], the Steering Committee did not reach consensus on a different risk adjustment alternative….The Steering Committee emphasizes that the lack of a specific alternative does not indicate assent with the Abt methodology.”

As NQF work on the final set of performance measures proceeded into early 2002, the NQF Steering Committee commissioned reviews from three outside experts to help with the complex, contentious issue of risk adjustment. All three agreed with the idea of making good information about nursing home quality available to the public and also that high standards for validity and technical quality were required of measures that would be interpreted as rankings of good versus less good nursing facilities. All three noted various technical problems with the measures developed by the Mega QI team, particularly the new risk adjustment method. One concern raised was that the method “over-adjusted” for risk and would inadvertently make some facilities look better than they were, as explained in a section below. Two of the three outside experts concluded that CMS should not publish QM using the Abt risk adjustment method.

One expert concluded, “The NQF Steering Committee and the research teams involved in developing these quality indicators and the risk adjustment approaches should be commended for the work they have conducted over the last number of months. However, based on my review of the information provided, I believe it would be irresponsible to report any of these nursing home quality indicators to the public without further study.”

A second expert concluded that CMS should “not use the [Abt risk adjustment method] as it is currently conceptualized….Its lack of validity…is likely to cause more problems than it solves.” The third expert arrived at a somewhat different conclusion, recommending among other things that CMS should “consider delaying the implementation of public report cards until

Zimmerman (author of the CHSRA QI). Additional members include representatives from provider organizations (e.g. GenesisHealth Ventures, a nursing home chain), consumer organizations (e.g., the National Citizens’ Coalition for Nursing Home Reform, AARP, and the Alzheimer’s Association), and purchasers (e.g., committee co-chair L. Partridge of the National Association of State Medicaid Directors).

76 Ibid.
adequate time has passed to address some of the unresolved methodological issues...[but] if it is not feasible to delay...use [the Abt risk adjustment method]....The consequences of publicly misclassifying facilities as poor facilities could be profound, particularly if the industry has concerns about the adjustment methodology used. It has been suggested that [the new method] is too conservative an approach....One could argue conversely that it behooves government to be conservative when it takes a step of this magnitude.”

CMS decided not to use Abt’s controversial new adjustment method for the pilot project. Instead, the eight selected measures were adjusted with two types of more conventional risk adjustment, explained below. CMS chose the measures for the pilot project from a longer list of recommendations proposed by NQF (Table 3.1). The pilot project was rolled out on April 24, 2002, with full-page advertisements in major newspapers placed by CMS and endorsements from the American Association of Homes and Services for the Aged (AAHSA) and the American Health Care Association (AHCA), the two major nursing home organizations, despite many concerns from some of their individual member facilities and state affiliates.

The original contract deadline specified that the NQF complete its full consensus process before CMS selected measures for reporting on all nursing homes in November 2002. As planned, the Steering Committee continued its deliberations, holding numerous conference calls and two additional public meetings. A draft report and preliminary set of core measures was posted on the Internet for public comment in March, then sent to NQF members for review on April 10. Notably, core measures in the March draft did not have the Abt risk adjustment method. A membership meeting to discuss the draft proposal was held on April 22, with member comments due three weeks later. Revisions to the draft were made based on comments received. Members submitted their votes on the measures to NQF during June and July. The final step in the process for NQF endorsement will be a vote by the NQF Board.

During NQF member voting on the measure set, CMS requested that the NQF Board of Directors delay final approval and endorsement until after the final validation report from Abt was available. The NQF Board decided to defer a final decision pending additional consideration in light of information that would become available during the late summer and fall of 2002: results of the pilot project, a final validation report from Abt, and a report of the California Health Care Foundation study of similar issues.

In June, AAHSA and AHCA sent a joint letter to CMS regarding the measures to be used in the November initiative. Although reaffirming their support of CMS’ public reporting initiative, these provider organizations expressed numerous concerns about their members’ experiences during the pilot project, the quality and timeliness of the data, and other matters. Their primary concern was appropriate risk adjustment: They urged CMS to use the new Abt risk adjustment method in the November national reporting, contending that failure to do so unfairly characterized certain types of facilities.

In early August 2002, shortly after the Abt research team completed its final validation report, CMS announced the ten QM it had selected for the November initiative (Table 3.1). The controversial new risk adjustment method, called a “facility admission profile” (FAP), was used for three indicators. In two of these instances (“residents with delirium” and “residents with pressure sores”), CMS reported QM with and without the FAP because both were found to be equally valid in the final study and because of “great stakeholder interest in how FAP affects

---

them. In the third instance ("residents who improved with walking"), only the version with the FAP was validated and hence was reported.

CMS explained that "As the process of choosing nursing home measures evolved, it became clear that the NQF needs adequate time to fully consider and discuss the Validation Report and to take into account any lessons learned from the Six-State Pilot Study. Once the NQF reaches consensus and delivers its recommendations, we expect to move to a timely implementation of the updated measures...." 82

### Table 3.1
Quality Indicators/Measures Selected by NQF (Provisionally) And CMS for Public Reporting

<table>
<thead>
<tr>
<th>Type of residents to whom indicator is applicable</th>
<th>NQF Steering Committee Recommendations for the 6 state pilot [*= those implemented by CMS (with slight modifications and some risk adjustment) in pilot]</th>
<th>NQF Preliminary/draft recommendations for the final core set</th>
<th>CMS Selections for the national initiative</th>
</tr>
</thead>
</table>
| Chronic Care                                      | 1. Late-loss ADL decline (needs more help doing daily activities)*  
2. Physical restraints*  
3. Pressure sores*  
4. Weight loss*  
5. Infections*  
6. Pain*  
7. Anti-psychotic drug use | 1. Late-loss ADL decline  
2. Physical restraints  
3. Pressure Sores  
4. Weight loss  
5. Pain  
6. Bowel or bladder incontinence  
7. Indwelling catheters  
9. Bedfast residents | 1. Late-loss ADL decline  
2. Physical restraints  
3. Pressure sores (with FAP adjustment)  
4. Pressure sores (no FAP) [weight loss omitted because it failed validation]  
5. Infections  
6. Pain |
| Post Acute (CMS: "those expected to stay a short time") | 1. Delerium*  
2. Pain*  
3. Improvement in walking  
4. Re-hospitalization during stay | 1. Delerium  
2. Pain  
3. Improvement in walking  
4. Improvement in walking (with FAP) | 1. Delerium (with FAP adjustment)  
2. Delirium (no FAP)  
3. Pain  
4. Improvement in walking (with FAP) |

### 3.2 KEY ISSUES

Three key issues related to the new CMS measures emerged during debates over the pilot project and throughout the NQF deliberations to date.

#### 3.2.1 The Scope of the Measures

CMS directed the Mega QI team to focus only on MDS-based measures. However, NQF advised CMS at an early stage that it believed it important to consider other performance indicators.

---

82 Ibid.
Many NQF members shared this view. For example, numerous reviewers of the NQF draft final recommendations urged inclusion of a staffing measure, citing a recently released, large-scale, federally sponsored study that found a strong relationship between nurse staffing and quality of care. CMS’ web site “Nursing Home Compare” currently includes staffing information, but evidence suggests that the accuracy of these data needs improvement. Researchers have found inconsistent relationships between quality of care (variously measured) and the staffing data currently on CMS’ web site; data from other sources (e.g., providers’ Medicaid cost reports) were used in the most recent large-scale federal study. Those data are difficult to obtain on a routine basis. In addition, appropriate staffing is related to resident care needs (case-mix), and disputes exist about whether a sufficiently good case-mix or risk adjustment method is available for a fair presentation of staffing data as a performance measure with the same status as the other new measures. The NQF will be reviewing these issues in light of members’ comments before issuing its final set of recommendations. As Table 3.1 shows, CMS did not include staffing as one of the final measures.

### 3.2.2 Risk Adjustment

Three types of risk adjustment are used in the new CMS measures. First, as CMS explains, “to reduce the chance that a nursing home that serves more frail residents is disadvantaged by the quality measures, information about some residents may be left out of the quality measures. For example, residents who move into a nursing home with pressure (bed) sores don’t count on the bed sore measure until after they have been in the nursing home for more than three months.” Second, all QM but one (“residents with restraints”) are statistically adjusted for resident clinical characteristics that predispose a resident to have a particular clinical problem (e.g., pressure sores). Third, three of the measures are presented with a special risk adjustment, the “facility admission profile” (FAP). As noted above, the FAP adjustment developed by the Mega QI team has been one of the most contentious issues.

A FAP is an adjustment factor applied to the calculation of a quality indicator at the facility level to address two issues that the Mega QI researchers believed would distort a facility’s quality rating: differential measurement practices (how well nursing home staff are able to complete MDS assessments) and differential selection (the types of residents the home admits). The researchers argued that homes should not be penalized in their quality ratings for these factors.

Failure to take into account differential measurement practices, the researchers argued, results in “ascertainment bias.” For example, the researchers note “Some facilities are especially attuned to psychological functioning of their residents. They may have on-site mental health professionals…who are involved heavily in care giving and training of staff. Such facilities may have higher rates of [reported] depressive symptoms among their residents and this does not signal a care problem.”

Failure to take into account differential admission practices, the researchers argued, unfairly penalizes facilities with QM that suffer from “selection bias.” For example, the researchers note “In the case of pressure ulcers, facilities care for residents with varying levels of intrinsic risk [that can be accounted for in traditional resident-level risk adjustment] such as

---

84 Comments are available at: http://www.NQF.org.
85 From the description of the QM used in the Pilot project: http://www.medicare.gov/NHCCompare. This type of risk adjustment is also used in the new QM.
functional problems, diabetes, incontinence, etc. Yet nursing facilities also inherit the results of the good and poor practices of hospitals….Since residents with a history of a pressure ulcer are significantly more likely to acquire one in the future, facilities admitting residents with pre-existing pressure ulcers run the risk of looking worse on a pressure ulcer QI simply because they admit a higher acuity population.”

The researchers hoped that a single FAP especially created for each QI (based on characteristics of residents admitted to a facility) would mitigate both ascertainment bias and selection bias, thus making the QI more accurate reflections of true quality of care differences among facilities. Preliminary testing of the FAPs in one state appeared promising to the researchers. For example, on the pressure ulcer QI, they found that homes with more skilled nursing staff had lower QI ratings without a FAP adjustment. Adding a FAP adjustment to the pressure ulcer QI brought facilities quality rankings in line with expectations from other research: the homes with more skilled nursing staff had better FAP-adjusted QI.

The three outside experts who reviewed this preliminary study agreed that the Abt team had identified some important issues but at least two were seriously concerned that the evidence in the preliminary report was not convincing. Among other things, concerns were raised that the new method “over-adjusted” the quality indicators. “Since the FAP (in the case of chronic care QI) is based on an admission assessment that might not be conducted for up to 14 days after admission, the FAP can actually reflect the early effects of care (both positive and negative) provided by the facility. During this delay, many of the quality concerns measured by the QI are likely to occur (that were not present upon arrival). Since these QI will show up on the assessment, the FAP will adjust…the facility’s QI rates (thereby making the facility look better than it should).”

The risk adjustment method was further tested in a large-scale validation study. In brief, the researchers found that the FAPs performed inconsistently. In most cases they appeared to make little difference; that is, QI with and without the FAP adjustment did equally well in identifying facilities that the researchers concluded were providing good care based on their assessment of care practices and other matters. In the cases where the FAP made a difference, half of the FAP-adjusted QI were more valid indicators of care quality and half were less valid. The researchers concluded “Overall, these comparative analyses demonstrate that, under the model used in our national validation of the quality indicators, there is little evidence that the FAP adjustment results in an array of QI scores that perform better than QI without FAP adjustment in the statistical sense. The FAP models did not, as had been hoped, out-perform the non FAP-adjusted models.”

The large-scale studies, sophisticated analyses, and intense scrutiny by NQF experts and members (including provider organizations) of risk-adjustment methods used in the new CMS quality measures support CMS’ claim that these measures represent scientific state-of-the-art. For some, however, the state-of-the-art is still too primitive for firm conclusions about comparative nursing facility performance.

87 K. Berg et al., Identification of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings, p. 147.
88 K. Berg et al., Identification of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings, p. 148.
89 J. Morris et al., Validation of Long-Term and Post-Acute Care Quality Indicators, p. 46.
90 Ibid. 48.
91 Ibid. 46.
3.2.3  Validity: To What Extent Do CMS’ New Measures Reflect Actual Differences in Nursing Facility Quality of Care?

Confidence in the validity of the CMS’ new QM rests largely on the results of the Mega QI team’s research. The challenge for the team was to develop a way to distinguish better performing nursing home from poor performing ones, independent of the QI they sought to test. To do so, they developed a set of “constructs” that they characterized as elements of good quality of care. Some of these were related to specific conditions; for example, “more frequent scheduling of assessments for suspicious skin areas.” Some were global ratings; for example, “the extent to which a facility manages clinical, psychosocial, and nutritional complications across domains in a manner consistent with high quality care.”92 Researchers collected data related to these constructs from interviews, observation, medical record review, and other sources in 209 nursing facilities in six states. Complex mathematical models were used to assess how strongly the selected QI were related to the quality constructs. The researchers concluded that the validation study provided “strong evidence that many of the…quality indicators capture meaningful aspects of nursing facility performance and are reliably measured.”93 They identified 6 chronic care and 4 post-acute care QI as those with the highest degree of evidence “that they represent real care processes with the highest level of validity.”94 CMS selected the new QM for national reporting from this list with one exception. The QM “residents with restraints” was judged only moderately valid, but has been retained for national reporting, in part because it was used in the six-state pilot project and has long been a focus of concern.

The extensive validity study substantially advances understanding about the new CMS measures and QI in general, although caution is still warranted regarding interpretation. One key question is the extent to which the quality constructs used to test the QI actually distinguish between good nursing facilities and those that perform less well. The researchers originally sought to develop a list of evidence-based procedures and processes known to reflect nursing home quality. However, they found that “there appear to be relatively few well-studied, research based ‘standards of care.”95 Thus, to a substantial extent, the validation criteria are based on face validity and expert opinion.

A second limitation is that the findings from this one study might not be more widely generalized to the entire U.S. population of nursing facilities. The study provides some evidence that findings are related to the sample of nursing facilities used in the QI testing. For example, the researchers found that the post-acute care indicator “improvement in walking” was highly valid in the sample of hospital-based facilities but was not valid when Medicare post-acute care patients from the freestanding facilities were added to the sample.96

92 The researchers began with a large number of quality constructs which they tested against QI. Only those constructs that were related in the expected direction to a sufficient number of QI were retained for the final analyses. The authors note that the reasoning here might be considered circular but “this is a process that characterizes most efforts at construct validation” (John Morris et. al., Validation of Long-Term and Post-Acute Care Quality Indicators, p. ES-4).
93 J. Morris et al., Validation of Long-Term and Post-Acute Quality Indicators, p. 60.
94 Ibid.
95 Ibid., p. ES-4.
96 Ibid., p. 52.
CONCLUSION

The first set of federally sponsored nursing home quality indicators was developed in the early 1990s. Since then, they have been refined and used for a broad range of purposes in nursing homes and related facilities. As this report has attempted to make clear, QI are pointers to potential quality problems. Any set of QI cover only certain aspects of quality. Care must be taken to select appropriate QI for the intended use, to understand the limitations of these measures, and be cautious in interpreting results.

Recent federally sponsored research discussed in this report has contributed substantially to the understanding about QI in general, providing far better support than previously existed for judicious use of selected measures in nursing home quality improvement efforts. Equally if not more important, has been the debate accompanying the ongoing National Quality Forum process. Well-qualified experts, steeped in the details of the new research, disagree about numerous important aspects of the new measures. The researchers themselves caution that the issues are highly complex and require further study. Thus, the matter of what constitutes “judicious use” remains open to debate.

Proponents of CMS’ new “Quality Measures” initiative argue that it will stimulate improvements in nursing home quality in two ways. First, if consumers use the performance measures to help select a nursing facility, poorer performing facilities will suffer in the marketplace and either improve or go out of business. Second, providers will work to improve performance on the measures (out of a sense of mission, shame, to attract more business, or avoid higher insurance premiums) and hence improve care quality. But if the QM misdirect consumers’, providers’, insurers’, and regulators’ attention from issues that matter more, then at best the QM will have a negligible effect on quality; at worst, consumers’ access to good quality care could suffer.

One lesson regarding the use of appropriate caution in interpreting the new QM has already been learned. Questions and answers (Q&As) posted on CMS’ web site about the measures used in the six-state pilot project included this question: “Are we confident that the performance measures accurately reflect the care furnished in each nursing home?” CMS answers: “Yes. Each performance measure measures an aspect of care related to nursing home resident outcomes. A national validation process now under way will link positive performance on each performance measure with good resident outcomes.” The “Yes” appears to have been premature in the case of one QI dealing with weight loss, which was found to be “not valid” in the Mega QI validation study. A different result, however, was found in the California HealthCare Foundation study that also examined the validity of the weight loss QI, but used a different research method. These differences provide further credence to the admonition that QI be used with caution and with a clear articulation of their purpose and audience.

The challenge for policymakers is to decide how much effort and resources should be devoted to working on improvements to nursing home QI, versus making investments in other activities related to improving and monitoring nursing facility performance. At present, little formal research is available to help them make those choices. At a minimum, however, sufficient resources need to be invested in QI explanation and analysis to counter potential misinterpretation of the data that could lead policymakers, stakeholders, and the public to draw erroneous conclusions about the meaning and implications of QI results.

USING THE CHSRA QUALITY INDICATORS IN NURSING HOME INSPECTIONS

TARGETING INSPECTIONS THROUGH THE NEW FEDERAL SURVEY PROCESS

In mid-1999, the federal government instituted a new nursing facility survey process that includes using the University of Wisconsin Center for Health Services Research and Analysis (CHSRA) QI. President Clinton’s announcement (July 21, 1998) of an initiative to improve the quality of nursing home care and subsequent Congressional hearings focused increased attention on the potential uses of QI as part of the solution, creating additional pressure for action. Despite widespread recognition that the CHSRA QI (and processes for ensuring fundamental data integrity) are works in process and that much remains to be learned about the potential effect on nursing home quality of carrying out a national survey and facility monitoring system that incorporates a fixed set of quantitative QI, the Health Care Financing Administration (HCFA) “view[ed] the need to move forward with using objective, resident level data in the survey process as too critical to wait until [research] to develop and/or evaluate quality of life indicators and to develop the MDS auditing protocol system [have been completed].” Training was begun before the system was completed. Provider organizations and some state Medicaid agencies expressed concern regarding surveyors being trained on and using a process that was not complete and was poorly understood, even conceptually, by most individual facilities. At the same time, HCFA had been under considerable pressure from Congress and some advocacy groups to move forward expeditiously with improvements to the nursing home quality assurance process. It remains to be seen whether the chosen course regarding implementation of the CHSRA QI proved to be an appropriate balance.

Researchers from the University of Wisconsin Center for Health Services Research and Analysis who developed the federally sponsored QI were also contracted by HCFA to develop

---

98 At least 2 of the President’s 10 point plan for improving nursing home care were directly related to QI: (1) “Implementing New Efforts to Measure and Monitor Nursing Home Quality: HCFA [is] collecting information on resident care through an automated data system that will be analyzed to identify potential areas of inadequate care in nursing homes and to assess performance in critical areas,” and (2) “Preventing Bed Sores, Dehydration, and Malnutrition: requiring state surveyors to monitor actions taken by nursing homes to prevent these ailments” (White House press release, July 21, 1998).


100 On April 19, 1999, a two-hour overview session was broadcast by satellite. In addition, “a four-day session for state survey directors and [HCFA] regional office officials [was] conducted during the week of April 26. HCFA [was also scheduled to give] three additional four-day training sessions for state and regional trainers between May 17 and 28” (www.ahca.org; “Regulatory Update,” May 1999).

101 Officially, final procedures are those specified in the Medicaid State Operations Manual (the “SOM”). In most cases, nursing home surveys are conducted by state surveyors who report to a state agency, which, in turn, has a contract with the federal government to fulfill that obligation. Thus, details regarding the nursing home survey process are specified officially in the SOM. A draft of the revised SOM was issued for public comment in February 1999. Final revisions were released in July 1999.
most of the procedures and training materials related to incorporating the CHSRA QI into the new survey process. CHSRA researchers caution that:

"Use of the QI and QI reports in the survey process offers an additional source of information from which surveyors or supervisory staff may make planning decisions about the survey of a facility and from which facility staff can plan their internal quality improvement initiatives. The QI and QI reports are not to be considered as a single source of information but should be used in conjunction with all pertinent information about a facility." 102

MDS data electronically submitted by each facility are used to create resident characteristic and QI profiles for each facility. Surveyors access facility profiles over the Internet as the starting point of the survey process. 103 Three types of facility reports are available. 104 Each report compares data for an individual facility to data for all facilities in that state.

- **The Facility Characteristics Report** provides information on the age, gender distribution of residents, payment source (e.g., the percentage of residents who are private pay or Medicaid), selected diagnostic characteristics (the percentage with a psychiatric diagnosis, mental retardation, or hospice care), 105 admission and resident “change in status” data, and additional selected information regarding residents’ assessed clinical stability and discharge potential.

- **The Facility Quality Indicator Profile** shows how the facility compares to others in the state with respect to each of the CHSRA QI. Among other things, this report “flags” QI when the facility exceeds an investigative threshold. Setting the threshold level for these flags has been (and is) a matter of some controversy. A facility will flag on this report under either of two conditions:
  - when the facility has even one case of a designated “sentinel health event” (current fecal impaction, dehydration, or pressure sores for someone at low risk); or
  - when the facility exceeds 75 percent of the facilities in the state with respect to a particular QI.

- **The Resident Level Summary** reports for each resident: name, assessment date and type, and check marks for any QI on which a resident “flags” (i.e., is identified as having a potential quality of care problem). For example, a resident identified as having a weight loss of 5 percent or more in the last 30 days or 10 percent or more in the last 6 months on the most recent assessment would be flagged on the weight loss QI.

The three reports are designed to be used in conjunction with other information (e.g., OSCAR and complaint data) to guide surveyors in selecting potential problem areas at a facility.

---

102 CHSRA, Nursing Home Quality Indicators National Data System, April 1999; emphasis added.
103 This technology is theoretically straightforward. As a practical matter, however, the initial training manual contains a full-page description of problems that likely would be encountered accessing facility profiles over the Internet. The manual notes that “Unfortunately, each version of each [Internet] browser has a slightly different behavior under otherwise identical conditions and despite our best efforts, we are left with a few bugs” (Ibid).
104 Ibid.
105 Nursing home care for residents with each of those characteristics has been the subject of particular concern regarding appropriate placement and appropriate use of Medicare and Medicaid program dollars.
and the residents to be included in the surveyors’ sample for actual review at the facility. Surveyors are directed to pay particular attention to “three areas highlighted during congressional hearings and oversight reports: pressure ulcers, weight loss, and dehydration.”

For example, surveyors are to select residents for on-site review disproportionately from among those with those conditions.

CMS has sponsored a study to evaluate the use of CHSRA QI in the nursing home survey. Results of the evaluation are expected in 2003.

SOUTH DAKOTA: PROPOSAL FOR COLLABORATIVE USE OF CHSRA QI IN AN ALTERNATIVE QUALITY ASSURANCE DEMONSTRATION

South Dakota was the first and to date the only state to request federal permission to test an alternative quality assurance system relying heavily on the CHSRA QI as measures of actual quality. As an early participant in CHSRA QI research, state officials perceived a need to translate quality measurements into improved quality outcomes. They saw the development of CHSRA QI as an opportunity to try to develop an alternative to the standard federal survey, with the goal of testing a different approach to quality assurance. One key aspect of the South Dakota approach was to harness the expertise of state surveyors to work in collaboration with individual facilities to design quality improvement strategies that would then be carried out by the facility. The project was called “The South Dakota Quality Initiative” (SDQI).

South Dakota submitted an application for a waiver to CMS (then known as HCFA) that would allow the demonstration to proceed. The waiver application was ultimately denied, but review of the proposed initiative and its demise is instructive.

The American Health Care Association (AHCA) was an active participant in SDQI from its inception. As noted in a recent federal report, “In 1996, at the request of the Governor, the South Dakota Department of Health, Office of Health Care Facilities Licensure and Certification and Department of Social Services began working with ACHA to develop ‘an alternative survey process for nursing facilities which would ensure compliance with regulations and improve the quality of care for nursing facility residents.’” Although it is both common and useful for public agencies working on projects that affect an industry to seek advice and feedback from industry representatives, the fact that AHCA became perceived as a prime mover in SDQI may have contributed to its ultimate defeat.

---

107 S. Nonemaker, CMS (formerly HCFA), personal communication, February 1999.
108 The South Dakota waiver application is discussed at some length in a report issued in 1998 by the U.S. DHHS and widely known as “the Deeming Report” (U.S. DHHS, Report to Congress: Study of Private Accreditation (Deeming) of Nursing Homes, Regulatory Incentives and Non-Regulatory Initiatives, and Effectiveness of the Survey and Certification System, 1998). The discussion above is based on three principal sources: (1) interviews with Joan Backman, the SD official with key responsibility for the project, The Deeming Report, and a close review of the actual SD waiver application.
109 U.S. DHHS, Report to Congress: Study of Private Accreditation (Deeming) of Nursing Homes, Regulatory Incentives and Non-Regulatory Initiatives, and Effect of the Survey and Certification System, Section 13.6.4.1 110 Testifying in 1999 before the IoM’s committee exploring ways to improve nursing home quality, American Health Care Association (AHCA) described its strong opposition to many aspects of the federal survey process and the shelving of the SDQI proposal. In the context of describing a need to move to an outcome-based quality measurement system, AHCA called for greater freedom for states to undertake demonstration projects (http://www.ahca.org/brief/quality.htm; November 1999).
The South Dakota initiative was based on the following principles:111

- The goal of a survey process must be improved quality.
- Quality in long-term care is measurable.
- Quality must be outcomes-focused.
- Quality must be customer-driven and customer defined...the resident is the primary customer.
- Quality must be measured from compatible data.
- Quality improvement must be a facility-internal process.

Elements of the program included a 5-year demonstration period for nursing facilities participating on a voluntary basis. Facilities were eligible if acceptable computer equipment and Continuous Quality Improvement (CQI) training for staff were in place. State officials estimated that two-thirds of South Dakota’s 116 facilities would be immediately eligible. The proposal called for initial participation of 80 facilities. Most South Dakota facilities have fewer than 100 beds, so the initial demonstration would have covered approximately 8000 beds.

The proposal called for the use of the then-current version of the CHSRA QI and customer satisfaction measurement (CSM) data from the AHCA Satisfaction Assessment Questionnaires. CSM data were to be collected using tools selected by the provider and approved by the state. Measures of customer satisfaction were to be taken at least every 90 days. Because South Dakota was a participant in the Multistate Nursing Home Case-Mix and Quality Demonstration, which demonstrated use of the CHSRA QI, the SDQI designers believed South Dakota to be uniquely qualified, not only to use the CHSRA QI competently, but to proceed to the next step—using the results to improve clinical outcomes.

The proposal included a detailed chart demonstrating how the CHSRA QI and an available CSM (Satisfaction Assessment Questionnaires of the American Health Care Association) and modified survey tasks could be combined to enable surveyors to evaluate compliance with each of the various requirements of OBRA 1987. The proposal also included a detailed evaluation program to be conducted by the various participants. In addition, provision was made for external evaluation.112

The SDQI proposal document requested such waivers as might be required to implement the program. Ultimately CMS (then known as HCFA) denied the request, officially citing lack of authority to grant relevant waivers. As mentioned above, state surveyors would then work with facilities to upgrade performance in areas in which they score low.

In discussing the SDQI, authors of an HHS study widely known as the “Deeming Report” note that, “In spite of the allusions to quality improvement and the inclusion of factors identified within the CMS [then HCFA] Quality Initiative Team's own definition of quality, the SDQI has

---


112 As noted in the proposal, “An independent expert will be retained by the State to monitor the implementation and evaluation of the demonstration. This expert will be asked on an annual basis to report on the impact of the proposed process on resident outcomes, resident and family member satisfaction, and the efficiency and effectiveness of the process...The State will provide written reports to HCFA on an annual basis with respect to its findings under the SDQI program” (Ibid).
not been popular with HCFA or national advocacy groups for several reasons,” citing the following:113

- “The partnering environment produced under the SDQI would reduce surveyor objectivity and the citation of noncompliance and perhaps lead to collusion…
- HCFA believed that the customer satisfaction measures that the SDQI would use were untested and unreliable…
- There was concern that State surveyors would become involved in the operation of the nursing facility and would take time away from resident care and other operational activities…
- QI cannot be used to measure quality of care or life or regulatory compliance since they are just now being validated and tested…
- South Dakota asked for permission to downsize the required resident sample for the annual survey of nursing homes that were in substantial compliance on the last survey
- South Dakota asked permission to provide technical assistance to NFs and skilled nursing facilities to incorporate the QI reports and the RAI into the facility continuous quality improvement activities…"

In all cases, South Dakota expressed a counter argument or a willingness to cooperate with HCFA in seeking common ground. Furthermore, authors of the federal Deeming Report described the project as “a reasonably developed proposal of a quality improvement initiative” and an initiative well-designed for program evaluation purposes.”114

However, the concerns raised by advocates and HCFA regarding relying too heavily on the CHSRA QI and customer satisfaction scores to gauge nursing home quality, rather than the federal survey process, led to HCFA’s rejection of the waiver.

In July 1999, the U.S. Secretary of Health and Human Services told a South Dakota audience that the federal Executive branch was reconsidering the matter of granting waivers for demonstration projects related to the survey process.115 Subsequently, however, HCFA officials stated that federal law precluded HCFA from granting waivers for any exceptions to the survey process where Medicare-certified beds were involved, although broader authority was available with respect to Medicaid.116 Those legal interpretations have been disputed. The new Administration that took office in January 2001 might interpret the laws differently or Congress might change them, opening the way for experiments in the future. Carefully designed and evaluated pilot projects could offer useful insights.

MAINE: USING THE CHSRA QI TO ASSESS TRENDS IN STATE NURSING HOME PERFORMANCE

Maine was one of the four primary states that participated in the National Case-Mix Demonstration during which the original CHSRA QI were developed. Along with the state

---

114 Ibid, section 13.7.2.
115 According to a reporter at the event, Secretary Shalala said “We would welcome the opportunity to review a written proposal once one has been jointly developed with the Department…” (N. Childs, “Providers Trapped in Survey Standoff,” Provider, May 2000, p. 24-38).
116 Ibid.
policymakers and staff who are actively engaged in nursing home policy development and implementation, the state benefits from close collaboration (under a special state contracting arrangement) with researchers at The University of Southern Maine’s Muskei Institute. These researchers, led by Professor Julie Fralich, developed and implemented innovative procedures for collecting and editing MDS data and producing facility-specific and statewide reports that have been used for several years in various ways. The Institute has developed software and reports that help the State with setting and paying Medicaid nursing home rates using a sophisticated case-mix system, helping quality assurance surveyors focus surveys, providing feedback to individual facilities with respect to QI, and conducting research on nursing home quality issues.

In Maine, state nursing home surveyors not only use QI reports in planning quality assurance survey visits, but they also bring laptop computers with access to many useful resident and facility-specific reports with them on each facility visit. State surveyors are responsible for assuring that facilities are appropriately assessing residents (using the federally mandated MDS) as part of the State’s process for insure the integrity of its case-mix payment system. Surveyors in Maine make more frequent visits than occur in many other states. According to Fralich, the mere fact that multiple official observers frequently visit facilities for various purposes is a factor perhaps as likely to enhance quality of care as the specific official tasks conducted during the visit.117

Just as it is difficult to interpret the meaning— with respect to actual quality of care of any one facility’s score on any one (or several) CHSRA QI —so too is it difficult to interpret trends in the aggregated data. Experienced researchers thoroughly exploring the initial year of data from Maine’s QI project concluded that, on balance, it appeared that performance on the QI was improving, but that it was still difficult to determine what that said about actual quality of care or to attribute changing performance to the introduction of a QI reporting system, rather than to other factors occurring at the same time.118

Table 1.1 below illustrates both the potential advantages of being able to track nursing facility performance using QI, and the pitfalls of a simplistic analysis. As can be seen, during the first year of project implementation, the prevalence of falls among residents increased by nearly 12 percent. Is that an indication of declining quality of care?

---

117 J. Fralich, personal communication, September 1998. Changed behavior resulting from the mere presence of observers (a common sense proposition) is known in social studies as “the Hawthorne effect,” after the name of a G.E. plant in which researchers conducted a study—over 50 years ago—of the effect of changing the lighting on workers’ productivity. The researchers found that—regardless of the level of the lighting—worker productivity was increased by the presence of the researchers. Numerous researchers, administrators of high quality nursing homes, and others have suggested over the years that “harnessing the Hawthorne effect” can be a useful approach to enhancing nursing home quality (for example, see: J. Barney, “Community Presence as a Key to Quality of Life in Nursing Homes,” American Journal of Public Health 64(2) (1974): 265-268.

118 Two important contemporaneous changes were (1) the state’s implementation of a case-mix payment system and (2) implementation of the new federal survey enforcement procedures, July 1995.
### Table 1.1
Change in Selected Quality Indicators: Maine: 12/94-10/95\(^{119}\)

<table>
<thead>
<tr>
<th>Quality Indicator</th>
<th>Percentage Change: All Residents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence of Daily Restraints</td>
<td>-16.4%</td>
</tr>
<tr>
<td>Prevalence of Falls</td>
<td>+11.8%</td>
</tr>
<tr>
<td>Prevalence of Any Injury</td>
<td>-20.7%</td>
</tr>
</tbody>
</table>

Although an increase in falls is not a good outcome, Maine found that the use of restraints had simultaneously decreased—a good outcome, although possibly associated with an increased prevalence of falls.\(^{120}\) But how serious were the reported falls? Perhaps increased attention to the accuracy of the facility resident assessments (associated with the State’s implementing a case-mix payment system) had led to the identification of more minor slips and falls. To explore this possibility, the researchers noted a simultaneous decrease in the prevalence of “any injury” (a QI derived from MDS data regarding fractures, abrasions, bruises, and burns).\(^{121}\) They found that performance on that QI had improved, indicating at a minimum another positive outcome on the face of it, and possibly suggesting that the increased prevalence of falls was not associated with increased actual harm to residents.

Additional factors, however, add complexity to interpreting the observed trends. For example, the increased prevalence of falls might be associated with more admissions of a sicker population (an anticipated consequence of implementing a case-mix payment system like Maine’s). This trend might mitigate one’s view of the relationship between increased falls and poorer care. Or perhaps the first observation (December 1994) and the last (October 1995) did not differ with respect to the proportion of people who fell and sustained a serious injury, but that relationship was obscured by a simultaneous decrease in minor bruises among people who had not fallen.

**CONCLUSION**

It bears repeating that QI are pointers to potential quality problems. Any set of QI cover only certain aspects of quality. Care must be taken to select appropriate QI for the intended use, to understand the limitations of these measures, and be cautious interpreting results.

This section has focused on two key uses of QI: as part of the nursing home quality assurance system and as an aid to monitoring aggregate nursing home performance. The new survey process incorporating CHSRA QI does focus attention on issues that have proved


\(^{120}\) The MDS User’s manual notes the following regarding the relationship between restraints and falls. “Although restraints have not been shown to safeguard residents from injury, one of the most common reasons given by facilities for restraining residents is to prevent falls. In some instances, restraints have been reported to contribute to falls and injuries. Because of the complications associated with restraint use, many physicians and geriatric clinicians recommend exploring alternatives for preventing falls, such as treating health problems and making environmental modifications” *(Long Term Care Facility Resident Assessment Instrument User’s Manual, Official Replica Edition*, Natick MA: Elliot Press, 1995, RAP 18, p. 2).

\(^{121}\) This CHSRA QI is not one of the 26 version 6.3 CHSRA QI implemented as part of the new survey process. Some of the data required for the “prevalence of any injury” CHSRA QI are not available on the quarterly assessments that are used for federal CHSRA QI construction. A closely related CHSRA QI in Version 6.3 is “Incidence of new fractures.”
problematic. As noted in Chapter 2, the principal risk stemming from technical problems with the CHSRA QI when they are used for targeting nursing home inspections is that misleading data could lead to overlooking problems. That is, for those 24 items, some homes with actual problems in those specified areas will be missed, while others will be thought to have problems when none is found. In addition, if surveyors rely too heavily on the 24 CHSRA QI items, despite strong cautions, then other aspects of quality and the survey process may be given less attention than is needed. Further, monitoring QI cannot substitute for a quality assurance system, although incorporating QI into the monitoring system could focus additional attention on quality aspects, as appears to be the case in Maine.

Finally, information available to policymakers through federally mandated collection of nationally uniform QI provides one potentially useful tool for monitoring aspects of nursing home quality. But, as the example from Maine illustrates, understanding these data is not simple. Analysts with access to a data set that includes all MDS assessment items for a state’s entire nursing home population (preferably covering a number of years) can use sophisticated statistical techniques to help further understand trends, untangling the sorts of complexities discussed above. Even with that, policymakers will need to look at data beyond QI to understand what is happening with respect to nursing home performance.

The challenge for policymakers is to decide how much effort and resources should be devoted to working on improvements to nursing home QI, versus making investments in other activities related to improving and monitoring nursing facility performance. At present, little formal research is available to help policymakers make those choices. At a minimum, however, policymakers will need to invest sufficient resources in QI explanation and analysis to counter potential misinterpretation of the data that could lead policymakers and to make erroneous conclusions.
APPENDIX B

USING QUALITY INDICATORS FOR INTERNAL QUALITY IMPROVEMENT

INTRODUCTION

Improving quality in nursing homes, as in other organizations, requires some way to assess change. Objective measures are helpful if these measures are reliable, valid, and focus on aspects of quality that are important. But the mere existence of formula and data for constructing QI means little without practical tools for managing and using the information (e.g., information technologies such as computer programs and benchmarking systems). Further, real quality improvements require both information and the knowledge to use it appropriately in the context of a total organizational approach.

A PILOT PROJECT EXPLORING HOW FACILITIES MIGHT USE QI BENCHMARKING INFORMATION

The Quality Indicator Index and Education (QUIIX-Ed) project was initiated by The American Health Care Association (AHCA) in 1994 to test use of an early version of the CHSRA QI. The goal was to use the CHSRA QI to support Continuous Quality Improvement (CQI) activities in nine Mississippi nursing facilities. Data were collected between March 1995 and March 1996. The project was one of several AHCA initiatives designed to identify ways to either improve or replace the current federal regulatory approach to nursing home quality assurance. The project explored the feasibility and results of providing CHSRA QI benchmarking data to facilities for their use in “internal, facility-driven efforts to improve the quality of clinical and functional care.”  

Project Design

Mississippi was chosen as the site of the project, in large part because machine-readable resident assessment data were available from all facilities in the state. Mississippi was one of the participants in the National Case-Mix Demonstration. As such, the state used the original, more extensive version of the resident assessment instrument, the MDS+.

Nine facilities were selected to participate, based on their enthusiasm for the project and the researchers’ efforts to work with facilities representing a range of types with respect to such factors as size, location, and resident characteristics. At the outset of the project, participating facilities were about average (compared to other facilities in the state) with respect to quality, as indicated on the CHSRA QI analyzed. The project team, with input from the facilities regarding which CHSRA QI were potentially the most useful for a CQI process, selected 15 CHSRA QI from those available at the time.

123“At the outset of the project the nine QUIIX-Ed facilities scored significantly below the state average on 6 of the 15 measures, significantly above the average on 3 measures, and about the same on the remaining 6 measures” (Ibid).
Participating facilities were provided specially designed software to calculate QI from assessment data entered by the facility as part of its routine MDS+ resident assessment process. On a monthly basis, the facilities sent a disc with the QI information to AHCA researchers who then generated reports that were sent back to each facility. The reports showed each facility how its performance on each of the 15 QI compared to the other 8 facilities and to statewide averages.

**Results**

Reported study results were based on three data collection and analytic activities: a survey of participants after the experiment; an analysis of how the QI were used in facility continuing quality improvement activities, based on on-site observations and 2 case studies; and quantitative analysis of QI performance.

Analysis of the survey of staff at participating facilities revealed the following findings:124

- “Staff found the quality indicator information to be valid and relevant to their CQI efforts.”
- “Six of eight facilities responding indicated that the quality information was ‘quite’ or ‘extremely’ useful to them.”
- “Six facilities also considered the information to have improved quality in their facility ‘quite a bit’ or ‘a great amount’.”

In their report, the researchers illustrated how facilities used the CHSRA QI benchmarking information by describing case studies at two facilities. In general, the researchers observed, “the most common approach to quality was the use of a quality assurance team…that responded to care issues as they arose, either on the basis of poor regulatory compliance results or frequent or severe negative outcomes. Quality of care issues observed by staff or the nursing director were also addressed.”125 CHSRA QI benchmarking reports, in the researchers’ view, helped facilities better identify “opportunities for improvement,” the phrase used to describe substantially below par performance.

The third approach to examining study results involved a quantitative analysis of QI performance. The researchers found that performance (among all the facilities combined) significantly improved for six of the 15 measures, comparing performance in March 1995 to that in January 1996. In no case did performance significantly decrease.126

**Did the introduction of the intervention make the difference?**

That important question could not be answered with any degree of established scientific certainty, given the study design—which is not to fault the researchers, as the optimal approach (a randomized, controlled study) is rarely feasible in social studies.

Evidence from a recently available study suggests that the decline in restraint use (and perhaps other significant changes observed in the experiment) may have been due—at least in part—to the implementation of new federal quality standards enforcement procedures in July 1995. Using a sophisticated quasi-experimental design and data from across the nation on four important resident care outcomes (including the use of physical restraints), researchers have

124 Ibid.
125 Ibid.
126 Reminiscent of the situation discussed in a previous section regarding Maine’s experience, the QUIIX-Ed researchers reported “After showing initial improvement, prevalence of falls returned to its initial level. Although this pattern is troubling, the change occurred while restraint use and bedfast occurrence were reduced” (Ibid).
recently found that during the same time as the Mississippi experiment was being conducted: (1) the use of physical restraints decreased nationally, and (2) the evidence suggests that “the new enforcement regulation was effective in improving resident status outcomes.”

In sum, this project is best viewed as an important pilot project, rather than as research offering substantial evidence regarding the positive value on nursing home quality of the chosen intervention (i.e., distributing monthly comparative QI reports). The narrow scope (nine Mississippi facilities) and duration (1 year) of the project, as well as factors confounding interpretation of the quantitative analysis limit extrapolation of the results. Nevertheless, the project—as an innovative and early application of the CHSRA QI—provided useful information leading to refinements in the CHSRA QI and related benchmarking systems. The report from this project (published in 1996) remains one of the very few articles in the publicly available literature regarding an actual use of QI in nursing facilities.

TOOL FOR PROVIDERS TO USE MDS-BASED QI: INCENTIVES AND OPPORTUNITIES FOR OTHER BUSINESS DEVELOPMENTS

The American Association of Homes and Services for the Aged (AAHSA) and the AHCA have each developed MDS data analysis programs and other tools to help nursing homes use CHSRA QI in quality benchmarking and improvement activities. AAHSA’s “MDS-Based Quality Improvement System” (QIS) offers participants an opportunity to send their MDS data on disc for CHSRA QI analysis to the University of Wisconsin Center for Health Systems Research and Analysis. Benchmark data come from a database of 700 to 800 participating facilities. AHCA’s “Facilitator” system provides software for facilities to analyze their MDS data and the CHSRA QI. Providers may also purchase a separate component that provides benchmarks from other facilities on the CHSRA QI. The benchmark analysis is supplied by a company called LTCQ, Inc.

LTCQ, Inc. also offers its own set of products including “Q-Metrics-IAS.” That system is described in marketing materials as providing “77 quality indicators and 16 risk-adjusted,

128 According to product literature, “The MDS-Based Quality Improvement System works with existing MDS software that can output data in the HCFA draft format for electronic submission (which protects confidentiality)” (http://www.aahsa.org/public/4pagflyer.htm; November 2000).
129 QIS is offered through AAHSA local affiliates. In 1999, the initial cost was about $900 and about $600 in subsequent years (M. Reilly, Field Director, Western Region, AAHSA; personal communication, June 1999). In 2000, the cost was $1200 (http://www.aahsa.org/public/pathlink.htm; November 2000). From CHRSA’s perspective, “The CHRSA/AAHSA project is not a commercial venture. It started out and remains a research project” (personal communication to AARP, August 2000).
130 The cost for the Facilitator “Quality Manager Software” was $995/year for non-members and free for members in 2000 (http://www.ahca.org/facpg10.htm; November 2000). Participation in the benchmarking system cost $1350/year for non-members and $1080/year for members in 2000 (Ibid). Recently, however, AHCA’s Board voted that AHCA should seek to provide the Facilitator system free to all members because “performance measurement is so critical to [the] industry” (http://www.ahca.org/quality/facpg10.htm; March 2001).
131 The founders and principals of LTCQ, Inc. comprise a list of experts in the field, including two researchers involved in the development of the MDS and other work for HCFA to develop new modules for the MDS and evaluate aspects of its application in federal payment and quality monitoring systems (http://www.ltcq.com; March 2001).
incidence-based performance measures.” The 16 QI cover many of the same elements as the CHSRA QI; however, the former are a proprietary product, while the CRSRA QI are in the public domain. LTCQ, Inc. encourages participating facilities to submit all their MDS resident assessment files electronically to the company on the same schedule as these files are submitted to the government.

Tools like those described above are essential if nursing homes want to actually use much of the information homes are required to collect on residents. Systems that integrate financial and patient (clinical) information are also important. However, concerns also have been expressed about the accuracy of MDS data.

At least two of the QI analysis systems described above are certified as “Performance Measurement Systems” for use in a new feature of The Joint Commission for the Accreditation of Healthcare Organizations’ (JCAHO) accreditation program. Beginning in 1999, JCAHO incorporated “performance measurement systems” involving quality indicators into its accreditation program. Although traditional facility standards remain at the core of JCAHO’s approach, the “ORYX initiative” requires long-term care facilities that seek JCAHO accreditation to:

- Select and contract with an independent organization (“PMS Contractors”) for a Performance Measurement System approved by JCAHO;
- Select two to six performance measures (defined as “clinical indicators” or “outcome measures.”) JCAHO’s performance measures are thus “quality indicators,” as that term is used in this report); and
- Submit quarterly QI data to the approved independent PMS Contractor, which will in turn report the data to JCAHO.

JCAHO believes that allowing facilities choices with respect to both a PMS contractor and QI is a positive feature of the system in that it provides “flexibility for the accredited organization to select a system from an approved list of systems and measures relevant to its

132 Ibid.
133 The 16 performance measures cover some clinical elements not covered by the CHSRA QI (e.g., pain), and many of the same elements as the CHSRA QI (e.g., pressure ulcers and weight loss). Further, the 16 QI are reported to be all risk-adjusted and incidence-based, while only a few of the CHSRA QI used in the federal survey process are risk adjusted or incidence measures. Published literature on the validity of the measures used in the LTCQ system was not available as of March 2001.
134 For a description of data file transmission to LTCQ, see: http://www.ltcq.com; March 2001). The federal government requires that MDS assessment data be sent first to designated State agencies for edit checks; the States then transmit these data to a federal repository. Access to data in the federal repository is carefully controlled because of privacy concerns.
135 A GAO Report to Congressional Requesters, Federal Efforts to Monitor Resident Assessment Data Should Complement State Activities GAO-02-279 February 2002
136 CHSRA’s performance measurement product for facilities that seek Joint Commission on Accreditation of HealthCare Organizations (JCAHO) certification costs $2200 per year per facility (http://www.chsra.wisc.edu/CHSRA/PIP_ORYX/faqs.htm; March 2000). LTCQ’s performance measurements for JCAHO accreditation are sold only in conjunction with one or more of the company’s other MDS-based products. The lowest cost combined product is $1200 per quarter (http://www.ltcq.com/purchase.htm; March 2001).
specific patient population and services.\footnote{138} Although this type of flexibility does have some positive features, it also raises the potential for “gaming.” For example, a provider might only select indicators on which it was certain to perform well, rather than those more relevant to identifying areas needing improvement.

The JCAHO ORYX initiative has the potential to increase attention to objectively measuring elements of quality and perhaps improve both facility performance and the JCAHO accreditation system. At the same time, it increases the number of agents and organizations with access to MDS data, raising complex issues related to privacy and selling of other products based on the MDS data. The confidentiality of medical records like the MDS is a growing concern.\footnote{139}

Businesses seek access to MDS data because these data can be used to create profitable products for the health care industry. Data-based products can be used by insurers to manage or avoid risky populations, help target sales of supplies and pharmaceuticals, help investors evaluate companies, help providers improve revenues from various payers, as well as help nursing homes with internal quality improvement activities.

**Some Principles for Using QI Effectively in Facility Quality Improvement**

Most nursing homes now view tracking their own performance on the same CHSRA QI that federal surveyors use as critical—at a minimum to manage their “regulatory risk.” Using QI to actually improve quality is a much more complex undertaking.

Table 1.1 in the report highlights some key features of QI measurement systems that are useful for internal quality improvement efforts, including focusing on processes that need to be improved by identifying gaps between scientifically based protocols and actual practice and giving feedback to those involved from very frequent measurements of relevant QI.

A group of noted geriatricians and nursing home researchers has applied those principles to improving nursing facilities in areas such as incontinence.\footnote{140} The researchers describe the necessity of identifying effective work processes (what nursing staff should do) to achieve better outcomes (greater maintenance of continence). This research has intensely focused on the actual process of care in nursing homes. The researchers have developed specific interventions to reduce incontinence (through individualized, targeted toileting programs) and also to reduce contractures and functional declines associated with excessive use of restraints.\footnote{141} These researchers believe that most nursing homes—however well intended—are unable to meet actual quality standards (i.e., quality care as defined in OBRA ’87) because homes lack appropriate technologies or processes for achieving those standards:

\footnote{138}Ibid.
\footnote{139} S. Joslin, CMS (formerly HCFA), personal communication, March 2001. The MDS assessment instrument records detailed personal information about individuals residing in nursing facilities. Federal regulations (drawing authority from the 1974 Privacy Act) hold nursing homes accountable for maintaining the confidentiality of these records. State laws, which sometimes have additional requirements regarding matters such as special protections for information regarding mental health and HIV status, also apply to nursing homes’ handing of MDS data. It is not clear that all transmissions of MDS data to agents other than the federal and state governments meet current standards for privacy protection and proper notification of patients. The Healthcare Insurance Portability and Accountability Act (HIPPA) specifies new rules designed to better protect the security and confidentiality of medical records, including patients’ right to authorize release of information. Initial regulations require health care providers, insurers, and other agents must be in compliance by February 26, 2003.
\footnote{141} Schnelle, 1996; Oustlander, 1987; Schnelle, et. al., 1998.
[Problems] are at least partially due to barriers created by the regulatory process. Mandating care standards without providing corresponding intervention technologies and analysis of the resources required to implement and maintain these interventions creates unrealistic expectations of nursing home providers. Providers react to these pressures by emphasizing paper compliance and chart documentation, rather than management of daily care practices....

The researchers argue that focusing on outcomes by monitoring QI may be one appropriate aspect of improving nursing home quality. They argue, however, that even improved monitoring systems have little probability of improving actual quality, in the absence of intervention technologies (and maintenance systems) that are “cost-effective and …proven to affect the quality indicators” and actual resident outcomes. Because developing such tested technologies is expensive, they urge government to invest more in research centers dedicated to such activities. In addition, the researchers have modeled the staff time required to perform effective care processes; in conjunction with other researchers, they have found that most nursing homes lack sufficient nursing staff to get the job done.

Appropriate technologies for improving quality outcomes (and QI) and sufficient resources to hire and train staff are necessary but insufficient. Implementing a successful prompted voiding program, for example, requires management commitment, very frequent QI measurement, and the active engagement of nursing aides in the process. The researchers developed a helpful table, replicated here as Table 2.1, to highlight differences between “traditional nursing home management” and “total quality management” (TQM).

145 The 10 principles of TQM are: constancy of purpose, customer-mindedness, quality-mindedness, process-mindedness, employee-mindedness, management leadership, benchmarking, continuous improvement, and suppliers as purchasers. The federal Agency for Healthcare Research and Quality (AHRQ, formerly AHCPR) sponsored a useful study examining how successful healthcare organizations (principally focusing on hospitals, some with attached nursing facilities) used benchmarking and other TQM principles. A report on that project (called “QISS”) and much additional useful work on long term care quality and practice guidelines can be found at: http://www.ahcpr.gov.
Table 2.1
Differences Between Traditional Nursing Home Management and TQM Approach

<table>
<thead>
<tr>
<th></th>
<th>Traditional Nursing Home Management</th>
<th>TQM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of work sampling</td>
<td>Low (arbitrarily determined)</td>
<td>High (determined by variability of outcome)</td>
</tr>
<tr>
<td>Who samples outcomes</td>
<td>Facility quality assurance staff; state and federal inspectors</td>
<td>Workers involved in the work process</td>
</tr>
<tr>
<td>Customer feedback</td>
<td>No regular sampling; react to complaints</td>
<td>Routine repeated samples (prevent complaints)</td>
</tr>
<tr>
<td>Employee training</td>
<td>Classroom training</td>
<td>On-the-job training</td>
</tr>
<tr>
<td>Content of training</td>
<td>No training in outcome monitoring or work analysis</td>
<td>Specific training in outcome monitoring and analysis of work processes</td>
</tr>
<tr>
<td>Worker motivation</td>
<td>Reprimands, warnings, small merit increases; facility citation system</td>
<td>Worker participation in evaluating and changing work processes</td>
</tr>
</tbody>
</table>

In the TQM approach, nurse aides are given the tools they need to improve outcomes (e.g., they are taught how to do “prompted voiding”); they have access to constant feedback on outcomes; and they participate in designing work process improvements to make the program more successful.

THE WELLSPRING PROGRAM

The Wellspring Program provides an example of quality indicators used in the context of a total organizational effort to achieve higher quality, cost effectively. The program has been independently evaluated, and the results were promising.

In 1994, a community of formerly competitive Wisconsin not-for-profit facilities formed an alliance of nursing home in order to apply group collaboration to the challenge of learning to deliver the highest quality, resident focused service in a demonstrably cost-effective manner. The starting point would be the then-new work in quality indicators to quantify quality of care with the goal of continually raising results. Benefits would have to be translated into dollars and cents. Thus was born the Wellspring Alliance, with 11 members initially, composed of rural and urban facilities, ranging from 63 to 115 beds. The strategies developed in that Alliance have now become a program that other facilities may learn and carry out under a licensing agreement.

---

Wellspring’s Executive Director has said that the most important action was breaking the paradigm of top-down management and empowering the staff. Management had grown accustomed to managing operational details. Under the new paradigm, management and staff have equal authority and responsibility for designing and managing methods. So long as the results are good and the method conforms to law and regulations, management’s role is defined as coach, not boss. For example, staff is permanently assigned to teams and permanently assigned to residents, which in turn fosters collaboration among staff. As a feature of the collaboration, staff manages its own schedule and calls in replacements when necessary. In practical terms, schedule management becomes a system of team back up in which team members call on one another as necessary. For the resident, the result is stable staffing and far fewer falls and altercations, both important quality outcomes.

The Wellspring Alliance developed a training and implementation program organized around seven modules representing the domains of clinical care quality. These training modules incorporate best practice clinical protocols from evidence-based research, federal standards, information about the CHSRA QI and other QI measurement, and substantial information on management processes.

Each facility selects a Care Resource Team from all levels and specialties, and all shifts, including nurses, certified nursing assistants, dietitians, social workers, office staff, and plant operations staff. A key point is choosing informal leaders in the existing organization. Training begins with a two-day intensive seminar followed by a 6-month implementation period featuring significant interactive follow-up.

Considerable attention is devoted to teaching critical thinking and analytic thinking skills. Adult learning principles and instruction on personality and learning styles are included. Role play is a part of the program. Case studies are included. Principles and practice of change management in organizations are presented and discussed; these principles are covered in each of the seven modules. The program takes a minimum of 2 years and sometimes up to 5 years to implement.

An important part of program’s apparent success is likely to be found in the follow-up, frequent interaction of Care Resource Team members, interaction among the teams, and engagement and support of facility management. Management training sessions (3 days) covering methods and implementation are held. Progress discussions are a part of monthly meetings attended by CEO’s.

In 1999, the cost of training alone was $7,000 per module. For one of the original Alliance members—Good Shepherd Services, employing 175 people—the peak cost in a year was approximately $55,000 for the program, a very substantial sum for Good Shepherd. However, Alliance members believed the investment worth the time and money. Results (as yet to be independently evaluated) through 1999 pointed to cost savings from improvements such as reduced staff turnover and indications of improvements in care.

149 The seven modules are: physical assessment; incontinence care; fall prevention; skin care and pressure ulcer prevention; nutrition; restorative care; and behavior management.
**APPENDIX C**

**USING QUALITY INDICATORS IN CONSUMER INFORMATION SYSTEMS**

**INTRODUCTION**

Easily accessible information about nursing home quality is sought for many reasons. Families are looking for help choosing a nursing home. Consumer advocates, news organizations, and researchers want more information. As nursing home occupancy declines in most areas, providers with better products want that known and welcome the “gold seal” of independent ratings. Many providers and others hope to better harness market forces to improve nursing home quality—an approach that hinges substantially on customers having useful comparative information, particularly before choosing a nursing facility.

**SOME ISSUES IN THE DESIGN OF REPORT CARDS USING QI**

CMS provides facility-specific information regarding nursing home deficiencies and some resident characteristics through its “Nursing Home Compare” web site. CMS will be expanding this information to include the new measures discussed in Chapter 3. The research cited in Chapter 3 suggests that those concerned about helping consumers compare nursing facilities need to focus substantial efforts on developing ways to help consumers better understand and use QI information appropriately as it becomes more widely distributed with the federal government’s imprimatur.

The potential value of publicly available, facility-specific QI information is informed by research indicating that little use is made of available performance indicators for hospitals or health plans. For example, in 1997, researchers interviewed representatives of “thirty-three large employers that purchase [health insurance plans on behalf of] 1.8 million covered lives.” They found, in brief, “that employers are not always aware of clinical outcomes data and that the measures do meet their decision-making needs.” Similarly, researchers found in 2001 that among those corporate employee benefits specialists surveyed who do get quality care reports, “26 percent said that the data are ‘difficult to understand, and 42 percent said the implications of the data are unclear.’”

The authors of the 1997 study offer several suggestions for improving the usefulness of performance data for employers purchasing health plans that are applicable for nursing home consumers and advocates as well. The researchers stress the need to avoid information overload and urge the development of decision support tools.

Three ideas recommended by the researchers, and adapted here, might well serve as part of an action agenda for those (including foundations, governments, and consumer advocacy

---


organizations) interested in making facility-specific QI information useful to potential nursing home customers:

- **Target purchasers:** Report cards, say the researchers, should not be “generic.” What works for a Medicaid agency will not be exactly the right approach for an individual consumer; and the right information for a person and family seeking post-stroke specialty care, will not be the same as that for a person or family looking for help for a person with dementia.154

- **Summary measures:** As challenging as these are to produce, the effort should be made. The key task is to strike an appropriate balance between the need for less complex information and summary measures that are sufficiently informative to be viewed as trustworthy and valid by the people who matter in particular decisions. This suggests the potential benefits of competing summary measures, discussed in the following section.

- **Computer-aided decision tools:** The researchers note “purchasers need help integrating many different types of variables into their choices. Computer-aided decision tools that elicit preferences and values can assist purchasers in making decisions that reflect their goals. A good decision-support tool should reduce the information-processing burden by breaking down decisions into manageable steps and guiding decision-making through pre-selected criteria. Such tools bring data together in a way that ensures that vital pieces of information are not lost and are properly weighed in decisions.”155

A number of books, publicly sponsored web sites, and consumer guides provide information on choosing a nursing home, for example, checklists to use in observing a home. Multi-dimensional scorecards and computer-aided decision tools could be a useful addition.

**AN EXAMPLE OF A MULTI-DIMENSIONAL NURSING HOME SCORECARD ON THE WEB; SUGGESTIONS FOR ADDITIONAL TOOLS**

Texas has had notable problems maintaining the quality of its nursing homes.156 As part of efforts to improve the situation, the State developed a multi-dimensional nursing home scorecard for comparing facilities, easily accessible on the Internet.157 The scorecard, which became operational in December 1999, combines information from the MDS (these data are aggregated to form 32 CHSRA QI for each facility), standard survey information, and complaint investigation reports.

Each facility is scored on four scales: three sub-scales and a summary score. The scales were developed from analysis of Texas data. In some of the scales, scores incorporate weights for different factors (i.e., give more value to certain things than to others) based on subjective

154 Participants at the 1998 CMS (formerly HCFA) conference, “Making Quality Count” stressed this same advice (see the main body of this report).

155 Ibid, p. 179.

156 In 1999, analysis of MDS data by the State revealed that of the 50 percent of residents who were incontinent, 90 percent had no toileting plan, 50 percent were said to be “inactive,” and the polypharmacy rate was approximately 35 percent (Dr. Leslie Cortes, Texas Department of Human Services, personal communication, May 1999). Further, a report by the Democratic staff of the US House Government Reform Committee, based on an analysis of State Survey data, found that 55 percent of Texas facilities had deficiencies with the “potential to cause actual harm” (N. Childs, “Texas Facilities Short on Resources, Report Says,” Provider, December 2000, p. 101).

157 See: http://facilityquality.dhs.state.tx.us.
judgments.\textsuperscript{158} Each facility also receives a summary score—ranging from zero (worst) to 100 (best), based on the facility’s scores on the other sub-scales. Using the paradigm made familiar by \textit{Consumer’s Reports}, the scorecard displays ratings on some of the scales with a center dot, filled, half-filled, or empty of color. In addition, the Texas web site carefully cautions the user about the limitations of the scorecard and provides additional information about how to choose a nursing home. The final result is a product that makes considerable information available in a manner that is easy to understand. Some Texas providers, however, believe that the information presents an unfair picture, citing outdated data among their concerns.

Every scorecard will reflect a certain point of view. Those that rely on any one set of QI or survey information incorporate the assets, limitations, and perspectives of those data sets. Incorporating the customers’ view by adding information from residents and families can provide an additional perspective. However, summing all the information in a helpful global scoring system means that one must decide to use or not use a weighting system (giving some features more importance in a score). If weights are used, someone needs to decide on some basis what they will be; any choice will be subject to dispute, and no one weighting system could fully reflect an individual’s value preferences.

At least two types of additional consumer tools could be helpful. First, as noted above, decision tools that take individual preferences into account would be helpful. These might be modeled after the decision tools that financial planners use to help individuals evaluate their “risk tolerance” against their desire for financial reward in planning their portfolios. Those financial planning tools incorporate information about the statistical variability of “performance measures” for stocks, bonds, and cash investments. Similar decision tools incorporating information about the known/expected variability of nursing home QI would be very helpful—to consumers, providers, and surveyors. Developing these tools will require more research, but such research is essential.\textsuperscript{159}

Second, nursing home evaluations and scorecards that are frankly idiosyncratic would also be helpful. Competing scorecards would mitigate the limitations of any one product. For example, newspapers might run columns discussing individual nursing homes much as newspapers now do restaurants.\textsuperscript{160} Currently, when the news media pay attention to nursing facilities, the focus is generally on negatives and exposés. Even though exposés are warranted in many cases, the focus on just the negatives obscures the complete picture, makes it even more

\textsuperscript{158} Representatives of a wide set of interests—clinical, consumers, and providers—participated in the system’s development.

\textsuperscript{159} The following example of the need to understand expected variability and performance is given by one set of researchers prominent in the field. “This situation involved a dispute between a nursing home and a federal survey team. The survey team entered the nursing home at night and found 65 percent of the residents wet. The nursing home was accused of not changing the residents on a 2-hour schedule as required by federal guidelines. The fault for not adhering to this 2-hour regimen was laid directly on the supervisory nurses and nurses’ aides. The fallacy of this argument was highlighted by published data demonstrating that in 10 different nursing homes, 65 percent of the residents would be expected to be found wet when a 2-hour changing regimen \textit{was} being consistently fulfilled. Unfortunately, a 2-hour changing system, no matter how consistently it is implemented, will result in a variable percentage of residents being found wet on any one check.” The authors go on to explain how instituting a data-driven and TQM approach to continence can maximize residents’ continence and motivate workers, as discussed in a previous section (J.F. Schnelle and J.G. Ouslander, “Total Quality Management: Administrative and Clinical Applications in Nursing Homes”).

\textsuperscript{160} Notably, the \textit{Washingtonian}, a monthly magazine about Washington D.C, runs an annual issue about “the area’s best doctors,” but nothing about “the area’s best nursing facilities.” Similarly a national news magazine publishes annual issues on “the nation’s best hospitals,” but nothing on “the nation’s best nursing facilities.”
difficult for nursing facilities to attract good workers, and further isolates nursing facilities from a public that increasingly wants to avert its eyes from the whole situation. More focus on “five star facilities,” even if defined in an idiosyncratic way, could be helpful.
Appendix D

Using Quality Indicators to Target Public Payments

Introduction

As common sense suggests, nursing home quality to a considerable degree is a function of the money that nursing homes spend and how they choose to spend it. The amount of money that nursing homes have to spend is heavily dependent on public payment systems. In 2000, Medicaid was the primary source of payment for 48 percent of nursing home expenditures, and Medicare was the primary source for 10 percent. Twenty-seven percent of expenditures were paid out of pocket (including about eight percent [8%] private health insurance). Those averages understate the importance of public payments in a majority of nursing homes because they mask the fact that in most states, most of the private paying residents, who typically pay higher rates than Medicaid, tend to be concentrated in relatively few homes.

The dominance of a few payers (Medicaid and Medicare) is just one of the several reasons why the nursing home “market” does not work like markets for consumer goods such as televisions. In unfettered markets, consumers—placing great weight on individual value in the tradeoff between cost and quality—make choices that drive the market to equilibrium ideally with perfectly aligned supply and demand at the “right price.”

Many hope to harness market forces better to improve nursing home quality. Few involved, however, actually support all of the preconditions of unfettered markets, which would include tolerance for nursing facilities to open, close, and fail anywhere and everywhere, solely in response to market signals. On the other hand, virtually everyone agrees, in principle, that public payment systems should not leave quality out of the equation. The sections below discuss two approaches to tying public payments to different quality indicators.

Tying Payments to Nurse Staffing Levels or Expenditures

One of the best-researched indicators of quality in nursing homes is nurse staffing. Numerous studies have found that more nursing staff and more money spent on nursing is associated with better quality, variously measured or determined. Thus, one strategy for tying

---

161 CMS (formerly HCFA), National Health Accounts, 2000 National Health Expenditures.
163 For markets to work as in economists’ models, at least four conditions are required: multiple competing buyers and sellers; free entry and exit from the market; unregulated products and prices; and consumers’ ability to distinguish clearly among products with respect to both cost and quality, coupled with consumer’s facing the full effects of prices.
164 Refer to 2002 staffing report
165 For example, the IoM concluded “The preponderance of evidence from a number of studies using different types of quality measures has shown a positive relationship between nursing staff levels and quality of nursing home care” (G. Wunderlich, et al., Nursing Staff in Hospitals and Nursing Homes: Is it Adequate? Washington, D.C: National Academy Press, Institute of Medicine, 1996). A detailed analysis of some of the limitations of previous research and the results of new studies demonstrating that more nursing staff improve quality outcomes is found in:
public payments to quality is to pay higher rates for homes with more nurse staffing and to avoid payment systems that reward facilities for having inadequate nursing staff.

Nursing home payment systems that pay a price for care regardless of individual facility expenditures are referred to as “pricing systems.” Medicare’s new payment system for skilled nursing facilities (SNFs) is an “all inclusive pricing system.” Regardless of what SNFs actually spend on care, Medicare pays them from a “price list” that specifies 44 different prices for patients in each of 44 different levels of care. Each SNF is paid the same price for each type of patient, with some adjustments for differences among places with regard to the cost of inputs such as labor and utilities.

Pricing systems like Medicare’s encourage facilities to reduce expenditures because they can profit on the difference between the payment rate and expenditures. But reductions in expenditures can reflect either improved efficiency or reduced quality—spending less on food and nursing care that residents actually need. Thus, the key dilemma in rate setting is appropriately balancing incentives for both cost containment and quality. In theory, an entire nursing home system could achieve balanced objectives with a rate-setting system that has strong cost-containment incentives, balanced by an equally strong and effective quality assurance system. Many State Medicaid program officials have historically been concerned about the ability of even a well-functioning quality assurance system to counter-balance exceptionally strong cost containment rate-setting incentives. This concern, in substantial part, accounts for the fact that only a handful of state Medicaid programs use pricing systems for nursing homes.166

To achieve a balance between both cost containment and quality objectives, virtually all states use prospective payment systems, with payments based in part on individual facility expenditures, but also with various limits and incentives designed to constrain costs. During the late 1980s and throughout the 1990s, many states implemented rate-setting systems that placed stronger cost-containment incentives (e.g., paid flat rates) on the portion of rates less directly related to resident care (e.g., administration), and weaker cost-containment incentives (e.g., by limiting profit) on the portion of rates most directly related to care (e.g., nursing). This continues to be a popular payment model. For example, each one of the four states (South Dakota, Maine, Mississippi, and Kansas) that put in place their first Medicaid case-mix systems as participants in the National Case-Mix Payment Demonstration instituted just such “modified” cost-related case mix systems.167

Good evidence exists that the choice between a cost-related and a pricing system makes a difference with regard to quality of care. One study of the 50 state Medicaid payment policies, controlling for other important factors (such as differences in case-mix and local labor costs) found that “Nursing homes in states with [pricing systems, all other things being equal, had] about 30 percent lower staffing per resident day … than those in [other states].” The researchers

---

166 Only a few states rely heavily on a nearly pure flat rate (pricing) system. Texas, California, and Arkansas have long been the purest examples, but Arkansas and Texas have recently changed to “modified pricing systems,” discussed above. In Texas, legislators changed the system in 1999 after years of concerns about poor quality and high profits. Oklahoma, Louisiana, Oregon, and Nevada also have systems that are typically classified as pricing systems; the Oregon system is new. Maryland at one time had a pricing system, similar to Medicare’s for determining payments for patient-care related items; but the State changed the system after becoming concerned about the incentives in the system.

167 New York and Texas—the other two demonstration states—had already implemented Medicaid case-mix payment systems.
concluded that “further unlinking of rates from costs is likely to reduce care hours available to [nursing home residents].”

Another study based on a nationally representative sample of nursing homes and residents found that nursing homes in states with pricing systems had lower quality of care, measured by case-mix adjusted staff-to-patient ratios and selected outcome measures. The structure of the Medicaid payment system was more important to quality than the level (amount) of the payment. In addition, the researchers found “that the difference in RN staffing intensity between cost-based and flat-rate systems is greater” where more beds are available. Because nursing home occupancy is declining nationally, this study suggests that Medicare’s pricing system may have a greater potential for adverse effects on quality if current occupancy trends continue.

PAYMENTS FOR ACHIEVING RESIDENT OUTCOMES

Few examples of nursing home incentive payments based on QI have been either demonstrated or carried out. The best known example is the San Diego Experiment conducted in 36 nursing homes in San Diego in 1980-1983. Those results have been evaluated multiple times with opinion still divided on the success of the experiment. However attractive the concept of incentive payments for quality improvement, it is not clear that there necessarily should be a measurable relationship between incentive payments and quality outcomes for residents. Incentives are normally related to motivating specific behavior. In the case of quality improvement, lack of motivation may not be the key issue. Insufficient funds, a difficult labor market, and ignorance (even among the best informed) regarding how one might actually improve outcomes, are likely to be equally or more important issues as lack of motivation. Further, if the major portion of the payment system is otherwise structured (i.e., as a “pricing system”) to create strong incentives to profit by reducing resources devoted to care, it is difficult to see how an incentive bonus could have much positive effect.

CONCLUSION

The development of federally sponsored QI has rekindled interest in the incentive bonus idea. But the issue remains contentious, given difficulties interpreting differences among facilities’ actual quality, based on differences in performance on one or more QI (especially when comparing smaller facilities). Minnesota, for example, has devoted nearly 4 years of effort to the idea, without agreement on a suitable approach. Pilot tests and experiments with a strong evaluation component would be useful.


170 Deeming Report, Section 12.6.2. For arguments in favor of basing nursing home payments on outcomes as a desirable, ultimate goal, see: R.L. Kane, “Outcome-Based Payment: A New Beginning,” Health Progress (April 1986): 26-70.