
Development and Testing of Nursing Home Quality Indicators

David R. Zimmerman, Ph.D., Sarita L. Karon, Ph.D., Greg Arling, Ph.D., Brenda Ryther Clark, R.N., M.S., Ted Collins, R.Ph., Richard Ross, and François Sainfort, Ph.D.

In this article, the authors report on the development and testing of a set of indicators of quality of care in nursing homes, using resident-level assessment data. These quality indicators (QIs) have been developed to provide a foundation for both external and internal quality-assurance (QA) and quality-improvement activities. The authors describe the development of the QIs, discuss their nature and characteristics, address the development of a QI-based quality-monitoring system (QMS), report on a pilot test of the QIs and the system, comment on methodological and current QI validation efforts, and conclude by raising further research and development issues.

INTRODUCTION

During the past several years, researchers at the Center for Health Systems Research and Analysis (CHSRA), University of Wisconsin-Madison, have developed and tested a set of indicators of quality of care in nursing homes, using resident-level data from the Resident Assessment Instrument (RAI) (Morris et al., 1990, 1991). These QIs have been developed to provide a foundation for both external and internal QA and quality-improvement activities.

The work reported here was supported by the Health Care Financing Administration (HCFA) under Cooperative Agreement Number 18-C-99256/5-04 and Contract Number 500-94-0010-WI-1. David R. Zimmerman and François Sainfort are with CHSRA and the Department of Industrial Engineering, University of Wisconsin-Madison. Sarita L. Karon, Greg Arling, Brenda Ryther Clark, Ted Collins, and Richard Ross are with CHSRA, University of Wisconsin-Madison. The opinions expressed are those of the authors and do not necessarily reflect those of the University of Wisconsin or HCFA.

The development of the QIs is a result of two related developments in the field of nursing home QA. The first is the growing interest among health care professionals, consumers, policymakers, and advocates about issues related to the quality of care and quality of life of nursing home residents (Institute of Medicine, 1986; Lang et al., 1990). These interests and concerns are reflected in the report of the Institute of Medicine (1986) study and in the subsequent passage of the Omnibus Budget Reconciliation Act (OBRA) of 1987. Among the important provisions of OBRA 1987 is the requirement that a comprehensive assessment of all nursing home residents using the RAI be conducted periodically. The RAI consists of the Minimum Data Set (MDS) assessment form and the Resident Assessment Protocols (RAPs). The MDS includes information about a resident's physical functioning and cognitive, medical, emotional, and social status. The RAPs are corresponding care-planning tools used to help identify potential care issues (Morris et al., 1991).

The second development is the Multistate Nursing Home Case Mix and Quality demonstration funded by HCFA. There are two objectives of the demonstration: (1) to develop and implement both a case-mix classification system (using the resident assessment information) to serve as the basis for Medicaid and Medicare payment and a QMS to assess the impact of case-mix payment on quality, and (2) to provide better information to the nursing home survey process. Four States (Kansas,

Maine, Mississippi, and South Dakota) are participating in both the Medicaid and Medicare components of the demonstration, and two additional States (New York and Texas) are included in the Medicare payment component. All six States are participating in the quality component of the demonstration, led by CHSRA.

QI DEVELOPMENT

The cornerstones of the QMS, the QIs, are derived from items on the Minimum Data Set Plus (MDS+), an enhanced version of the MDS. Under OBRA 1987, the MDS was mandated for administration on all nursing home residents in the Nation. The MDS+ was developed to obtain additional data believed: (1) to be important measures of resource utilization (thus necessary for the development of a case-mix payment system), or (2) to have important implications for the measurement of quality of care. The MDS+ contains detailed information on a resident's physical and cognitive functional status, acute medical conditions, nutritional status, behavior, and emotional status. It also includes limited information on various processes of care, including a detailed inventory of current drugs being administered and the use of physical restraints. The MDS+ provides longitudinal resident-level data. Each resident in participating facilities is assessed when first admitted to a nursing home, each quarter thereafter, and whenever there is a significant change in functional or health status. Additionally, residents who are transferred to a hospital for treatment of an acute problem are assessed upon readmission.

The QIs were developed through a systematic process involving extensive interdisciplinary clinical input, empirical analyses, and field testing. Clinical and research staff at the University of Wisconsin-

Madison developed an initial draft of a set of indicators and potential associated risk factors based on an extensive review of relevant clinical research and the care-planning guidelines from the RAPs. The initial draft was then reviewed by several national clinical panels representing the major disciplines involved in the provision of nursing home care (including nursing, medicine, pharmacy, medical records, social work, dietetics, physical, occupational, and speech therapy, as well as resident advocates and administrators). The clinical panels provided a rigorous critique and assisted in refining or deleting proposed QIs and defining new ones. The clinical review culminated in the panels being convened in July 1991 to provide an assessment of the QIs within and across disciplines. This important step was then followed by an indepth review by a research advisory panel convened to provide consultation in areas of analytic concern. The panel members have continued to provide consultation throughout the project. The result of the clinical panel meeting was a set of 175 QIs organized into the following 12 care domains:

- Accidents.
- Behavioral and emotional patterns.
- Clinical management.
- Cognitive functioning.
- Elimination and continence.
- Infection control.
- Nutrition and eating.
- Physical functioning.
- Psychotropic drug use.
- Quality of life.
- Sensory function and communication.
- Skin care.

These 175 QIs have served as the basis for empirical analyses. QI development has been guided by several criteria, including clinical validity, feasibility or usefulness of the information, and empirical analyses.

Extensive analyses have been performed to further reduce the set of QIs to a comprehensive set of useful indicators. We have continued to revise the QIs through empirical testing and field review. One part of this review process was the QI pilot test, described later. The final set of QIs to be used in the quality component of the demonstration is similar but not identical to the set used in the pilot studies. Revisions in the QIs have been made on the basis of the following factors:

Results of the pilot tests—The feedback and quantitative analysis of the pilot test findings were instrumental in making revisions to the QI definitions and the system for incorporating their use into the proposed demonstration survey process. For example, surveyor difficulty in using differing denominators when interpreting facility QI rates resulted in a decision to use more prevalence QIs and fewer incidence QIs. Pilot test feedback was also instrumental in changes in the report formats.

Empirical analysis—Analysis of the data from the four Medicaid-Medicare States has continued since the beginning of the development process. This analysis has been instrumental in defining the QIs and risk factors, determining which types of MDS+ assessments should be included in the identification of potential care problems using the QIs, and establishing relative and absolute standards for use in targeting facilities and problem areas. For example, QIs that involve multiple MDS+ items to construct an index or scale have been made more parsimonious by identifying items with low prevalence or that are highly correlated with other items. These issues are discussed in more detail later.

Clinical input—Clinical input has been solicited on both a formal and informal basis throughout the development and testing process. This input has been essential in establishing the face validity of deci-

sions resulting from the empirical analysis and, in several cases, has had an important role in reversing those decisions. This is especially true in cases where low-prevalence items were recommended for exclusion from a QI definition, but where the clinical conclusion was that the validity of the QI would be questioned without the item, despite its low prevalence.

CHARACTERISTICS OF QIs

The QIs are markers that indicate either the presence or absence of potentially poor care practices or outcomes. QIs represent the first known systematic attempt to longitudinally record the clinical and psychosocial profiles of nursing home residents in a standardized, relatively inexpensive, and regular manner by requiring the expertise of only inhouse staff.

The QIs can best be described by looking at their characteristics from three perspectives: (1) resident versus facility level, (2) prevalence versus incidence, and (3) process versus outcome.

Resident Versus Facility Level

At the resident level, QIs are defined as either the presence or absence of a condition. The resident-level QIs can be aggregated across all residents in a facility to define facility-level QIs. These can then be used to compare any given facility with another or with nursing home population norms at the State or multistate level. An example of a resident-level QI is the prevalence of (stage 1-4) pressure ulcers, defined as "one" if the resident had such ulcers on the most recent assessment and "zero" otherwise. The corresponding facility-level indicator is the proportion of residents of a facility that have one or more pressure ulcers—that is, the number of residents with pressure ulcers on the most recent assessment, divided by the total

number of residents in that facility. At both resident and facility levels, several QIs have associated risk factors. These are health or functional conditions that either increase or decrease the resident's probability of having a specific QI. For example, the factors defining high risk for the prevalence of pressure ulcers are: impaired transfer or bed mobility, hemiplegia, quadriplegia, coma, malnutrition, peripheral vascular disease, history of pressure ulcers, desensitized skin, terminal prognosis, diabetes, and pitting edema. A resident who has one or more of these conditions is believed to have a higher likelihood of having one or more pressure ulcers, as indicated by the QI. Risk factors also are used to adjust for interfacility variation in QI scores.

Prevalence Versus Incidence

At both the resident and facility levels, a QI that is defined as the presence or absence of a condition at a single point in time is called a "prevalence QI," whereas a QI capturing the development of a condition over time (on two consecutive assessments, for example) is called an "incidence QI." It should be noted that, although prevalence QIs relate to a single point in time for each resident, at the facility level they represent the prevalence of conditions over a 3-month period, because the most recent assessment across the population of residents can occur over a quarter.

Process Versus Outcome

QIs cover both process and outcome measures of quality. Donabedian (1980) describes quality of care as ". . . that kind of care which is expected to maximize an inclusive measure of patient [or resident] welfare, after one has taken account of the balance of expected gains and losses that attend the process of care in all its parts." To fully

measure quality of care requires a complete accounting of the interplay between and among structural, process, and outcome measures. Process indicators represent the content, actions, and procedures invoked by the provider in response to the assessed condition of the resident. Process quality includes those activities that go on within and between health professionals and residents. Outcome measures represent the results of the applied processes. Outcomes refer to the ". . . change in current or future health status that can be attributed to antecedent health care" (Donabedian, 1980). In the case of long-term care, it maybe more relevant to think in terms of a change in or continuation of health status. Outcome quality, then, would include questions of how the resident fared as a consequence of the provision of care, i.e., whether the resident improved, remained the same, or declined. Hence, outcome indicators should be represented by both point prevalence and incidence measures.

The distinction between a process and outcome QI is not always straightforward. The distinction can be addressed along two dimensions. In some cases, the QI is a combination of an outcome and a process, in that it reflects both of them. An example is the presence of symptoms of depression with no treatment indicated. In these combination cases, we identify the QI as being both an outcome and process measure. In other cases, the QI can be considered either an outcome or a process measure, as illustrated by the QI "little or no activity." This variable can be considered to reflect the status of the resident (i.e., the resident is not able to or chooses not to engage in activities) or a process QI (i.e., the facility staff elects not to provide or arrange for the activities). In these cases, we have chosen the conservative approach and considered the QI an outcome measure. Subsequent investigation, of course, may determine that, for a

Table 1
Quality Indicators and Risk Adjustment Used in Demonstration Facility and Resident Reports

Domain	Quality Indicator	Type of Indicator	Risk Adjustment
Accidents	Prevalence of any injury	Outcome	No
	Prevalence of falls	Outcome	No
Behavioral and Emotional Patterns	Prevalence of problem behavior toward others	Outcome	Yes
	Prevalence of symptoms of depression	Outcome	No
	Prevalence of symptoms of depression with no treatment	Both	No
Clinical Management	Use of 9 or more scheduled medications	Process	No
Cognitive Patterns	Incidence of cognitive impairment	Outcome	No
Elimination and Continence	Prevalence of bladder or bowel incontinence	Outcome	Yes
	Prevalence of occasional bladder or bowel incontinence without a toileting plan	Both	No
	Prevalence of indwelling catheters	Process	Yes
	Prevalence of fecal impaction	Outcome	No
Infection Control	Prevalence of urinary tract infections	Outcome	No
	Prevalence of antibiotic or anti-infective use	Process	No
Nutrition and Eating	Prevalence of weight loss	Outcome	No
	Prevalence of tube feeding	Process	No
	Prevalence of dehydration	Outcome	No
Physical Functioning	Prevalence of bedfast residents	Outcome	No
	Incidence of decline in late-loss activities of daily living	Outcome	Yes
	Incidence of contractures	Outcome	Yes
	Lack of training or skill practice or range of motion for mobility-dependent residents	Both	No
Psychotropic Drug Use	Prevalence of antipsychotic use in the absence of psychotic and related conditions	Process	Yes
	Prevalence of antipsychotic daily dose in excess of surveyor guidelines	Process	No
	Prevalence of antianxiety or hypnotic drug use	Process	No
	Prevalence of hypnotic drug use on a scheduled or as-needed basis greater than twice in last week	Process	No
	Prevalence of use of any long-acting benzodiazepine	Process	No
Quality of Life	Prevalence of daily physical restraints	Process	No
	Prevalence of little or no activity	Outcome	No
Sensory Function and Communication	Lack of corrective action for sensory or communication problems	Both	No
Skin Care	Prevalence of stage 1-4 pressure ulcers	Outcome	Yes
	Insulin-dependent diabetes with no foot care	Both	No

NOTE: Late-loss activities of daily living are bed mobility, eating, toileting, and transfer.

SOURCE: Zimmerman et al., Center for Health Systems Research and Analysis, University of Wisconsin-Madison, 1995.

particular resident, the QI is more reflective of a process of care than of resident status.

The QIs were designed to cover both process and outcome of care and to include both prevalence and incidence types of measure.

A set of 30 QIs, covering all 12 domains, has been selected for use in the QMS in the multistate demonstration. These QIs have been selected on the basis of empirical analysis, clinical review, and the results of

the pilot test (described later). The 30 selected QIs are presented in Table 1, which also classifies each QI as process, outcome, or both, and notes whether the QI has associated risk factors.

Development of the QI-Based Quality Monitoring System

Concurrent with the refinement of the QIs, we developed a system for using them

Table 2
Excerpts From a Facility-Level Report on Nursing Home Quality Indicators

Domain or Quality Indicator	Residents With Quality Indicator	Residents in Denominator	Facility Proportion	State Proportion	Percentile Rank
Accidents (Domain 1)					
Prevalence of Injuries	9	73	12.3	19.7	26
Prevalence of Falls:					
High Risk	0	61	0.0	15.0	0
Low Risk	0	12	0.0	8.1	0
Behavioral or Emotional (Domain 2)					
Problem Behavior:					
High Risk	5	39	12.8	35.9	6
Low Risk	4	34	11.8	10.5	55
Symptoms of Depression	7	66	10.6	8.9	66
Elimination and Continence (Domain 5)					
Incidence of Bowel or Bladder Incontinence:					
High Risk	1	26	3.8	14.0	11
Low Risk	1	27	3.7	5.3	43
Bowel or Bladder Incontinence Without Toileting Plan	17	22	77.3	36.5	96
Incidence of Indwelling Catheters	0	65	0.0	0.0	0
Prevalence of Fecal Impaction	0	73	0.0	0.4	0

SOURCE: Zimmerman et al., *Center for Health Systems Research and Analysis, University of Wisconsin-Madison, 1995.*

as a source of information in the survey process. We began with site visits to each of the four demonstration States to observe a survey and meet with survey staff, facility staff, residents, and industry representatives. Since that time, we have closely monitored developments in the survey process at both the national and State levels to ensure that the QI-based system is consistent with that process and to facilitate the integration of the QIs into it.

The QIs are used in the survey process to identify areas of potential concern. This information can be used by survey teams prior to and during the survey to identify areas that may warrant special focus and to identify residents that may be good candidates for inclusion in the indepth sample. To facilitate the use of the QIs, we developed reports, designed to present both facility-and resident-level information, for use by State project staff. The purpose of the facility-level report is to provide an overview of the QIs in each facility by presenting the prevalence and incidence of the QIs, as well as a comparison to peer group (State) averages. A quick review of this report can highlight issues that may

be of concern, for example, if a facility is well above the State average in the prevalence of pressure ulcers. The facility report also presents the percentile rank of the facility, in relation to its peers. Facility ranks are provided for the facility as a whole, unadjusted for risk, and separately for the QI occurrence among high- and low-risk residents, respectively. Although this approach does not provide a basis for overall comparison of the facilities adjusted for case-mix differences, it has the advantage of highlighting risk-related issues for the surveyors, who may have unique concerns for each of the distinct risk-based groups. These data facilitate the selection of QIs for review, based on their relative rankings. An example of a facility-level report is presented in Table 2.

The resident-level report provides information about individual resident's conditions and care practices as defined by the QIs. This report is structured as a matrix, in which the residents are indicated in the rows and the QIs are indicated in the columns. This report format allows quick identification of all the residents in a facility who have a particular QI, as well as iden-

tification of the full range of QI issues experienced by any individual resident. An example of a resident-level report is presented in Table 3.

The second component of the QI-based QMS is a series of protocols covering each of the QI domains. These protocols are intended to serve as guides to the surveyors using the QIs as the basis for investigating the adequacy of care in a facility. The protocols incorporate information from the current HCFA (1992) *State Operations Manual* (SOM), used by surveyors in conducting their survey field visits and making determinations about regulatory compliance and adequacy of care. The SOM information is rearranged to facilitate the integration of the QIs into the overall survey process and to permit the surveyor to use a QI as the basis of the survey information gathering task when a potential quality problem has been identified through that QI.

PILOT TEST OF THE QI SYSTEM

A pilot test of the use of the QIs and reports in the survey process was conducted in 1993–94. The pilot tests were conducted using 31 QIs covering 11 of the 12 domains. A total of 32 pilot test surveys were conducted in 3 States (Maine, Mississippi, and South Dakota). The primary objective of the pilot tests was to assess the feasibility and utility of the QIs and reports in the regular survey process and to assess the accuracy of the data items comprising the QIs. A secondary objective was to obtain preliminary information on the validity of the QIs in accurately identifying care problems at the resident and facility levels. In each of the participating States, designated surveyors used the QI reports and protocols in selected surveys. Training sessions were held with each of the designated surveyors to familiarize them with

the QIs, the facility- and resident-level reports, and the feedback mechanisms developed as part of the pilot tests. Manuals were provided to the designated surveyors for use in the field survey visits. The nursing homes chosen to receive pilot surveys were selected on the basis of convenience and scheduling considerations.

Information from the pilot surveys was collected in several ways: surveyor feedback forms, return of resident-level QI reports with surveyor notes on them, telephone debriefing calls, and review of the formal statement of deficiencies. Surveyors completed feedback forms for 14 pilot surveys in South Dakota, 10 pilot surveys in Mississippi, and 8 pilot surveys in Maine. The feedback forms elicited surveyors' documentation and opinions on the utility of the QIs in identifying potential problem areas, resident selection, and decisionmaking. A system was developed for indicating on the resident QI reports the accuracy of the QIs, the residents that were reviewed as part of the survey, and the linkage, if any, between the QI and a quality-of-care problem. Debriefing telephone calls were used to elicit further information on the feasibility and utility of the QIs for the pilot surveyors. These calls provided the surveyors and case-mix pilot coordinators opportunity to review problems and concerns, have questions answered, receive support from the CHSRA research staff, and offer recommendations that they could not easily communicate on the feedback forms. The calls also allowed the CHSRA researchers to have a fuller understanding of the surveyors' experiences in using the QIs in this way. The case-mix pilot coordinators collected the materials from the surveyors and mailed them to CHSRA staff, along with the final statement of deficiencies.

The results of the pilot tests were encouraging and useful in refining the design of the demonstration's quality com-

Table 3
Excerpts From Quality Indicator (QI) Resident-Level Summary Report

Resident Name	Resident Identification Number	Date of Assessment	Assessment Type ¹	Resident Age	Resident Gender	QI ²	
						Prevalence of Injuries	Prevalence of Falls
Resident A	01	4/13/95	A	86	F	—	—
Resident B	02	5/10/95	O	63	M	None	—
Resident C	03	5/27/95	Q	95	F	—	—
Resident D	04	7/07/95	Q	75	F	—	—
Resident E	05	6/21/95	Q	76	F	—	—
Resident F	06	5/05/95	Q	54	F	—	—
Resident G	07	6/18/95	A	85	F	—	—
Resident H	08	7/13/95	Q	93	F	—	—
Resident I	09	5/05/95	Q	93	F	—	—
Resident J	10	7/07/95	O	91	M	—	—
Resident K	11	7/14/95	Q	90	F	—	—
Resident L	12	6/24/95	Q	92	F	—	—

Resident Name	Problem Behavior	Prevalence of Depression	Use of 9 or More Medications	Cognitive Impairment	Prevalence of Cognitive Impairment	Incidence of Decline in Cognitive Status	Incidence of Bowel or Bladder Incontinence	Incidence of Bowel and Bladder Incontinence Without Toileting Plan	QI ²
Resident B	—	—	—	—	—	—	—	—	
Resident C	—	—	—	None	—	—	—	—	
Resident D	—	—	None	—	—	—	—	—	
Resident E	High	—	—	None	—	—	—	High	
Resident F	—	—	None	—	—	—	—	Low	
Resident G	—	—	—	—	—	—	—	—	
Resident H	High	None	—	None	—	—	—	High	
Resident I	—	—	—	—	—	—	—	—	
Resident J	—	—	—	—	—	—	—	—	
Resident K	High	—	—	None	—	—	—	High	
Resident L	—	—	—	None	—	—	—	High	

¹ A = annual, Q = quarterly, and O = other.

² Table column entries indicate that the particular QI was flagged or not flagged for that resident. Entries indicate that the QI was flagged, as follows: none = QI is not risk adjusted; high = resident is at high risk for the QI; low = resident is at low risk for the QI. A dash indicates that the QI was not flagged.

SOURCE: Zimmerman et al., Center for Health Systems Research and Analysis, University of Madison—Wisconsin, 1995.

ponent. With respect to feasibility, the surveyors in general found the QI reports easy to interpret and integrate into their presurvey and survey activities. Most found that reviewing the reports was not disruptive to their normal survey functions. A few surveyors had difficulty interpreting the facility comparison information, particularly the use of differing denominators in facility proportions or rates. This led to refinement in the definition of some QIs, a subject discussed more fully in the next section. Disruption also resulted when facility staff had inaccurately coded the MDS+ items, but often this provided useful information in identifying noncompliance with the resident assessment regulations.

The surveyors also found, in general, that the QIs were useful in helping to focus the survey activities, including the presurvey review, facility tour, selection of in-depth sample residents, and the quality-of-care and quality-of-life aspects of the information gathering tasks. More than 80 percent of the comments made by surveyors expressed a positive experience with using the QIs as part of the survey. Some surveyors noted that the QIs were useful as a basis for focusing the selection of residents and gathering information—even in cases where ultimately there was not a finding of deficient care—in part because it provided more confidence that the issue had been adequately addressed and the conclusion was valid. Many helpful comments were received about how the reports and protocols could be improved to facilitate the use of the QIs.

As already noted, the surveyors also assessed the accuracy of the MDS+ items comprising the QIs. First, they determined whether the information on the QI report was consistent with the information on the actual MDS+ completed by the facility staff member. They also determined whether

there was evidence in the resident's record supporting the entry on the MDS+, or in the absence of such evidence, that at least there was not evidence to the contrary. The surveyors were looking for both systematic programming errors and for MDS+ interpretation and completion errors on the part of facility staff. The QIs were determined to be accurate when both the MDS+ and other information from the resident's record supported the QI definition. The pilot study accuracy findings for selected QIs are presented in Table 4.

In general, the QIs were found to be accurate in the pilot tests, with the vast majority of QIs exhibiting accuracy rates above 85 percent. In only two cases across all three States did the pilot tests identify what were potentially systematic problems with accuracy. In each of these cases, we worked with the State data analysts to identify and correct what were computer coding errors. All other errors were idiosyncratic and could be traced to data-entry errors. An important cautionary note is that the determination of QI accuracy was not the primary objective of the pilot tests, and they were not designed to provide definitive conclusions about accuracy. For example, the facility sample design, as noted earlier, was selected on the basis of convenience and survey scheduling considerations and is not necessarily representative of the facility population in the three States. Similarly, the QI problem areas and residents were not selected to be representative of their respective populations; rather, the surveyors were instructed to use the QIs, including the selection of QIs and residents, in a manner that would maximize consistency with their existing survey procedures. A more rigorous evaluation of the accuracy and the validity of the QIs, described in a subsequent section, is currently underway as part of the demonstration technical assistance activities.

Table 4
Results of Pilot Test Investigation of Quality Indicator Accuracy

Quality Indicator	Risk Group	Number of Cases Investigated	Percent of Cases Accurate ¹
Prevalence of Any Injury	No	26	100
Prevalence of Falls	No	53	96
Prevalence of Problem Behavior	All	47	98
	High	35	97
	Low	10	100
Prevalence of Symptoms of Depression	No	35	100
Prevalence of Use of 9 or More Scheduled Medications	No	36	100
Prevalence of Cognitive Impairment	No	83	98
Incidence of Decline in Cognitive Status	No	14	79
Incidence of Bladder or Bowel Incontinence	All	36	89
	High	24	96
	Low	5	80
Prevalence of Incontinence Without a Toileting Plan	All	48	85
Prevalence of Fecal Impaction	No	8	100
Incidence of Indwelling Catheters	All	6	100
Prevalence of Urinary Tract Infection	No	28	97
Prevalence of Antibiotic or Anti-Infective Use	No	39	97
Prevalence of Weight Loss	No	40	93
Prevalence of Tube Feeding	No	8	88
Prevalence of Bedfast Residents	No	17	88
Incidence of Decline in Late-Loss Activities of Daily Living	All	37	97
	High	27	96
	Low	10	100
Incidence of Improvement in Late-Loss Activities of Daily Living	All	18	83
	High	12	84
	Low	3	100
Incidence of Contractures	All	29	97
	High	8	100
	Low	12	100
Incidence of Decline in Late-Loss Activities of Daily Living Among Unimpaired Residents	No	27	100
	All	38	74
Prevalence of Antipsychotic Use	High	16	62
	Low	18	78
	All	18	100
Incidence of Antipsychotic Use Following Admission or Readmission	No	1	100
Prevalence of Antipsychotics Exceeding Guidelines	No	8	100
Prevalence of Anti-Anxiety or Hypnotic Drugs	No	25	96
Prevalence of Hypnotic Use on a Scheduled or As-Needed Basis Greater Than Twice in Last Week	No	14	100
	No	62	100
Prevalence of Long-Acting Benzodiazepine Use	No	40	98
Prevalence of Daily Physical Restraints	No	17	94
Prevalence of Little or No Activity	High	13	100
	Low	1	100
Incidence of Pressure Ulcer Development	All	11	100
Prevalence of Diabetes Without Foot Care	No	3	100

¹ Cases in which the investigator found the quality indicator to be accurate.

NOTES: No = not risk adjusted; all = both high- and low-risk residents included; high = only high-risk residents included; low = only low-risk residents included. Late-loss activities of daily living are bed mobility, eating, toileting, and transfer.

SOURCE: Zimmerman et al., Center for Health Systems Research and Analysis, University of Wisconsin-Madison, 1995.

USE OF QIs IN THE DEMONSTRATION

On the basis of the empirical analyses, clinical panel input, and the pilot tests, a QMS to be used in the multistate demonstration has been developed. The QIs are

the heart of the system, which has been designed to incorporate them into an experimental version of the current survey process (which underwent extensive changes in summer 1994). Although some elements of this QI-driven QMS are specific to the Federal survey process, many of

its components have general applicability to any external or internal nursing home QA or quality-improvement initiative.

The demonstration QMS will use the QIs in the following ways:

Identifying facilities—In combination with other measures (previous deficiencies, complaints, etc.), the QIs will be used to identify facilities that may have more serious or particular types of care problems, on the basis of a comparison with their peers. These facilities may be subject to more extensive and/or more frequent QA monitoring. In this way, the QIs can be used to vary the monitoring process such that the intensity and frequency of the process can be more commensurate with the likelihood that care problems will be found.

Identifying areas of care—The QIs will also be used, again through peer comparison, to identify particular areas of care that might warrant a more indepth review during the survey or other type of visit. They can also be used to identify areas in which the facility appears to have no specific problems relative to other facilities or to some absolute standard. In either case, the initial indications can be confirmed through onsite observation by the surveyor or monitor. The previsit QI review by care area can also help identify special monitoring resources that might be required. For example, if potential medication-related problems are identified, a pharmacist could be added to the monitoring team. Because specialized clinicians, such as pharmacists or dietitians, are typically scarce resources, the ability to deploy them more appropriately can make the monitoring process more efficient as well as more effective.

Identifying residents—Because QIs and reports are defined at the resident level, they will also be used to identify residents who are good candidates for inclusion in the indepth sample, for more detailed

review in the monitoring process. Each identified resident can be confirmed or replaced with another resident in the sample on the basis of the onsite review by the surveyor or monitor. The resident-level reports will also provide supplemental information from the MDS+ that can assist the surveyor or monitor in gathering information needed to determine whether adequate care is being provided.

Structuring the collection of information—Through the use of the aforementioned protocols, the QMS will facilitate the gathering of information by the surveyor or monitor by using the QI presence as the basis for the investigation. This is expected to provide a more efficient framework for information gathering and decisionmaking.

Followup monitoring—The QI reports will be used to monitor progress or recurrence of problems following the survey and formal followup activities. QI reports will be run periodically for each facility, and previous (or new) problem areas will be reviewed. The reports will prompt additional visits if necessary.

Data-driven complaints—Another proposed feature of the system is the use of QI reports to identify problem areas that might develop between surveys or monitoring visits. Using more stringent standards to identify more serious problems, the QIs can be used as the basis for special monitoring activity, similar to the filing of complaints from residents, family members, or other advocates.

In combination, these features will permit the testing of a system that is based on the existing (revised) Federal survey process, but with the enhancement of taking advantage of resident- and facility-level assessment information to facilitate the planning and operation of the QMS.

Use of the QIs for purposes of quality monitoring will require high levels of data accuracy. The fact that the same (MDS+)

data are being used for both payment and quality monitoring helps to ensure the quality of the data. Incentives to over-report higher case-mix residents will be balanced by the fact that such misreporting will also increase the likelihood of being identified as an outlier on some QIs. Similarly, the tendency to under-report items indicating potential quality-of-care problems will have the result of reducing case mix and, therefore, payment. Additionally, regulatory requirements and enforcement related to data accuracy will remain in effect. Thus, the quality of data collected from the demonstration should be no worse, and may be better, than data collected outside of the demonstration.

METHODOLOGICAL ISSUES

In developing the QIs and the method for using them, we have been faced with several methodological challenges. Each of these challenges can be met in several ways. It has been our task to determine the most appropriate method, given the goals of the work. We recognize that other approaches could have been chosen. Indeed, we also might choose other approaches, particularly in cases where the QIs will be used for purposes outside of the survey process. The following sections describe the primary methodological challenges and our current approaches to each.

Assessment Type

Within the demonstration, resident assessment data must be collected at several points in time: at initial admission to the facility; quarterly after admission; annually (fourth quarter) after admission; upon significant change in health or functional status; and at readmission from a hospital (or other treatment facility). A small number of assessments also are performed for other reasons. For instance, in South

Dakota, assessments are conducted at 30 days after the initial admission. It should be noted that the distinction between quarterly and annual assessments is trivial, as each annual assessment is also obviously a quarterly assessment. Quarterly assessments currently are not required outside of the demonstration, although the new version (MDS 2.0) of the RAI will require completion of a partial MDS on a quarterly basis. Thus, tracking the annual assessments as distinct from the quarterly assessments allows us to assess the data that would be available under the current situation outside of the demonstration States.

Information collected at these different times has different relationships to the care provided in the facility. Information collected at initial admission provides baseline information but does not represent outcomes or processes of care provided in the facility. Information collected at readmission can represent outcomes of care provided outside the facility, or outcomes of care within that facility that may have resulted in the need for hospitalization. For these assessments, the relationship between a QI and the quality of care provided by the facility is unclear. Information collected both at admission and at readmission also can provide insight into a facility's admission practices. Assessments conducted quarterly (annually) or upon significant change in resident status can be more safely assumed to reflect the quality of care provided by facility staff.

Given these differences among the reasons for assessments, it is not surprising that the occurrence of a QI can vary in frequency by assessment type. An example of such differences is provided in Table 5. The prevalence of falls in the last 30 days is lowest (7 percent) among residents whose most recent assessment was simply a routine quarterly assessment. The prevalence of falls for residents whose most recent

Table 5
Prevalence of Three Quality Indicators, by Type of Assessment

Assessment Type	Number	Prevalence of:		
		Falls	Antipsychotic Drug Use	Pressure Ulcers
All Assessment Types	38,709	11.1	Percent 11.2	13.6
Initial Admission	6,068	16.3	8.1	25.4
Readmission	4,146	23.7	10.3	19.5
Significant Change	1,936	20.4	11.0	24.5
Quarterly or Annual	25,222	7.0	12.3	8.8
Other	1,337	11.7	6.6	17.9
Assessments Included in Facility Comparison ¹	28,495	8.1	11.9	10.3

¹ Assessments included in facility comparisons are significant change, quarterly or annual, and other. Initial admission and readmission assessments are not included in facility comparisons.

SOURCE: Zimmerman et al., Center for Health Systems Research and Analysis, University of Wisconsin-Madison, 1995.

assessment was for an initial admission is more than twice that rate (16 percent). For these residents, a recent fall may have necessitated their admission to the facility. The rate of falls is even greater (24 percent) for residents whose most recent assessment is for readmission. For these residents, we cannot tell whether the fall occurred in the facility, perhaps necessitating the hospital admission, or whether the fall occurred during the time that the resident was hospitalized. Thus, we cannot attribute the fall to the quality of care provided by the facility. We can note, however, that the rate of falls indicated on significant change assessments (20 percent) is nearly as high as the rate for readmissions. Other patterns of differences in QI rates among assessment types are also presented in Table 5. The prevalence of antipsychotic drug use, for example, is greatest among those in the quarterly/annual assessment group; but the differences between quarterly assessments, readmission assessments, and significant-change assessments, are quite small.

In measuring the prevalence of a QI for a facility then we have had to consider whether or not to include data from all types of resident assessments. The issue of which assessments to include in interfacil-

ity comparisons is a controversial one. In this early stage of QI development, we have elected to take the conservative approach of including in the comparisons only those assessment types that unambiguously reflect care provided in the facility. This means that, for purposes of interfacility comparison, we have chosen to calculate facility-level QIs on the basis of quarterly, annual, or significant-change assessments. Interfacility comparisons also include a small number of assessments defined as "other." These represent special cases, sometimes defined by the individual States, as in South Dakota. By excluding readmission assessment data from our calculations of the QIs, we increase the likelihood that the potential problem captured by the QI is rooted in care provided within the facility. The tradeoff is that we may miss cases of poor care that result in conditions necessitating treatment in a hospital (i.e., false negative). It is our belief that, where poor care does result in need for hospitalization, we also will observe its consequences among residents who experience the same condition but do not require hospitalization for treatment.

The effects of this decision are illustrated in Table 5. The first and last lines respectively show the prevalence of the example

QIs when all assessment types are considered together and when only the assessment types that we have elected to use for interfacility comparison are considered together. As seen in the table, the choice of which assessments to consider can have a sizable impact on the QI prevalence rates. For instance, the prevalence of falls is 11.1 percent when all assessments are considered but only 8.1 percent when using the select group of QIs. The prevalence of antipsychotic drug use, on the other hand, is slightly greater when considering only the select group of assessments (11.9 percent) than when considering all assessment types (11.2 percent).

Identification of Risk Factors

A second methodological challenge has been the consideration of risk factors. This challenge has required us to distinguish between two very important uses for risk information: (1) the identification of clinical risk factors, to facilitate the provision of appropriate and high-quality care, and (2) the identification of differences in facility populations that could result in different rates of QI occurrence where there is no difference in the quality of care provided.

The RAPs are an example of the first use, in that they identify factors that place the resident at higher risk of experiencing an adverse event or condition. For instance, persons who have impaired bed mobility may be at increased risk of developing pressure ulcers. It is critical that facility staff be able to identify these risk factors in order to provide appropriate care. This statement implies that the presence of a risk factor does not necessarily result in an adverse outcome. Indeed, where facilities are providing high-quality care, there may be no such increased risk. Therefore, it is important to identify risk factors for the purpose of establishing fair comparisons among facili-

ties. In doing so, however, it is not necessarily appropriate to include factors that the facility staff can use to both identify increased risk and to intervene in some way to prevent the higher risk from translating into a higher probability of the adverse event or condition.

In developing a system to permit interfacility comparisons of quality and the identification of facilities with potential quality-of-care problems, we have attempted to avoid using risk factors that are directly related to the quality of care. A system of risk adjustment for purposes of measurement of facility quality must exclude, as much as possible, the use of risk factors that the facility can reasonably be expected to identify and treat to avoid the outcome of the QI. Risk factors used for facility comparison must instead focus on issues that differentiate the populations, but where the ability of the facility staff to intervene is believed to be minimal. This concept can be expressed in the following way:

$$\text{QI} = \text{quality of care} + \text{risk} + \text{error}.$$

Given this purpose for the QIs, we have revised the original, RAPs-driven set of risk factors, to focus on those issues that we believe are not easily amenable to clinical interventions. Stated another way, we have attempted to develop a set of "pure" risk factors, such that:

$$\text{Risk} = \text{QI} - \text{quality of care} - \text{error}.$$

Risk-Adjustment Procedures

Implicit in the preceding discussion is the idea that risk-adjustment factors can be used to "level the playing field" when comparing quality across facilities. There are several ways in which risk can be taken into account. Once again, the purpose for which the QIs are intended must be a

major consideration in settling on an approach. One method for adjusting risk is to use standard epidemiological methods to create a single risk-adjusted rate for each QI for a facility. Using this approach, one would consider the expected rate of QI occurrence, given the presence of various risk factors. The ratio between the observed rate of occurrence and the expected rate of occurrence would provide a measure of quality. Facilities in which this ratio exceeded 1.0 could be assumed to have a potential problem with the quality of care, whereas facilities for which the ratio fell below 1.0 could be assumed to be providing exemplary care.

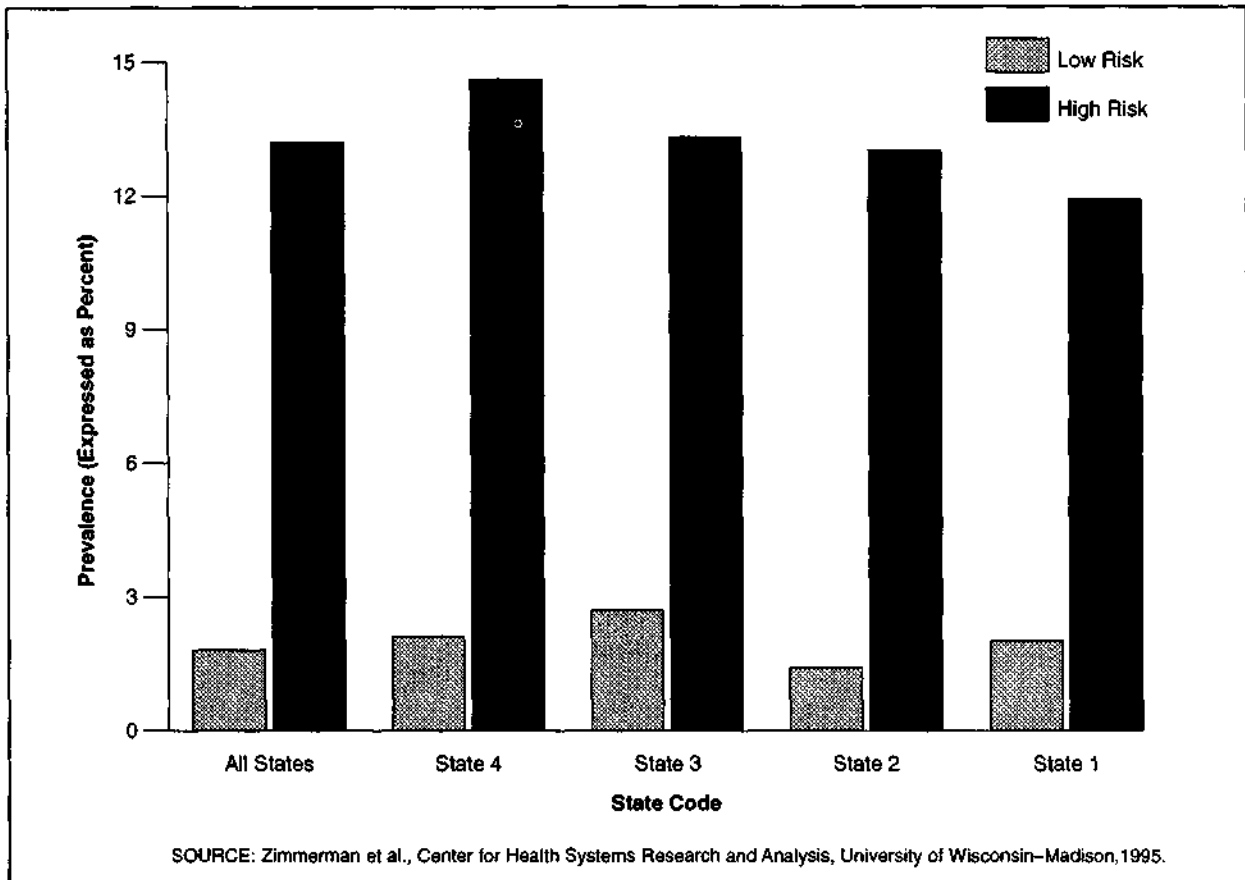
For purposes of the current demonstration, we have chosen a different approach. The epidemiological approach can be very useful in research and in making global comparisons. However, we believe it is less useful in the survey process, where surveyors may be more interested in the detail that goes into calculating such a number than in that number itself. Therefore, we have used a more direct method for applying risk. For each QI that has a risk-adjustment factor, we have created what are essentially three separate measures. The first is the occurrence of the QI without regard to risk. The second and third measures are the QIs measured separately for those people who have the risk factors and for those who do not. For the sake of simplicity, we have referred to these groups as "high risk" and "low risk," respectively. By creating separate measures for the populations defined by risk, the surveyors can: (1) determine the relative sizes of the high- and low-risk populations for a facility; (2) identify whether the facility has a potential quality-of-care problem for either or both risk groups; and (3) identify whether the facility has a potential quality-of-care problem for the resident population as a whole. Figure 1 shows the prevalence of pressure ulcers by level of risk.

This approach also allows us to set separate thresholds for the high- and low-risk groups. This can be important if we believe that the occurrence of a problem is more or less acceptable in these different groups. For instance, we may be willing to accept some (low) level of occurrence of pressure ulcers among the high-risk group. On the other hand, our tolerance for pressure ulcers among the low-risk group may be much less. We may believe that the occurrence of pressure ulcers among this group is much more likely to be an indication of a problem with the quality of care.

Quality Performance Standards

Using the QIs as a mechanism for identifying facilities where there is a potential quality-of-care problem requires a method for defining when such a problem is likely. This is accomplished by setting a standard or threshold for performance above which a facility's performance is considered suspect. There are several approaches to setting such thresholds. At the most basic level, thresholds can be either absolute or relative. Absolute thresholds can be developed based on review of the literature or on a consensus of the experts. These standards may be as low as zero, so that any occurrence of a QI signals a potential problem for the facility. Sometimes these cases are called "sentinel events" in QA parlance. Relative thresholds are peer-group based. They can be set at a level based on the distribution of the events across facilities, e.g., the 75th percentile, the 90th percentile, the mean plus two standard deviations (i.e., the 95th percentile, in a normally distributed population). Regardless of how the threshold is determined, it has implications for the cost and resources required of the survey process. The lower the threshold, the greater the number of facilities that will be identified for review.

Figure 1
Prevalence of Pressure Ulcers, by Level of Risk

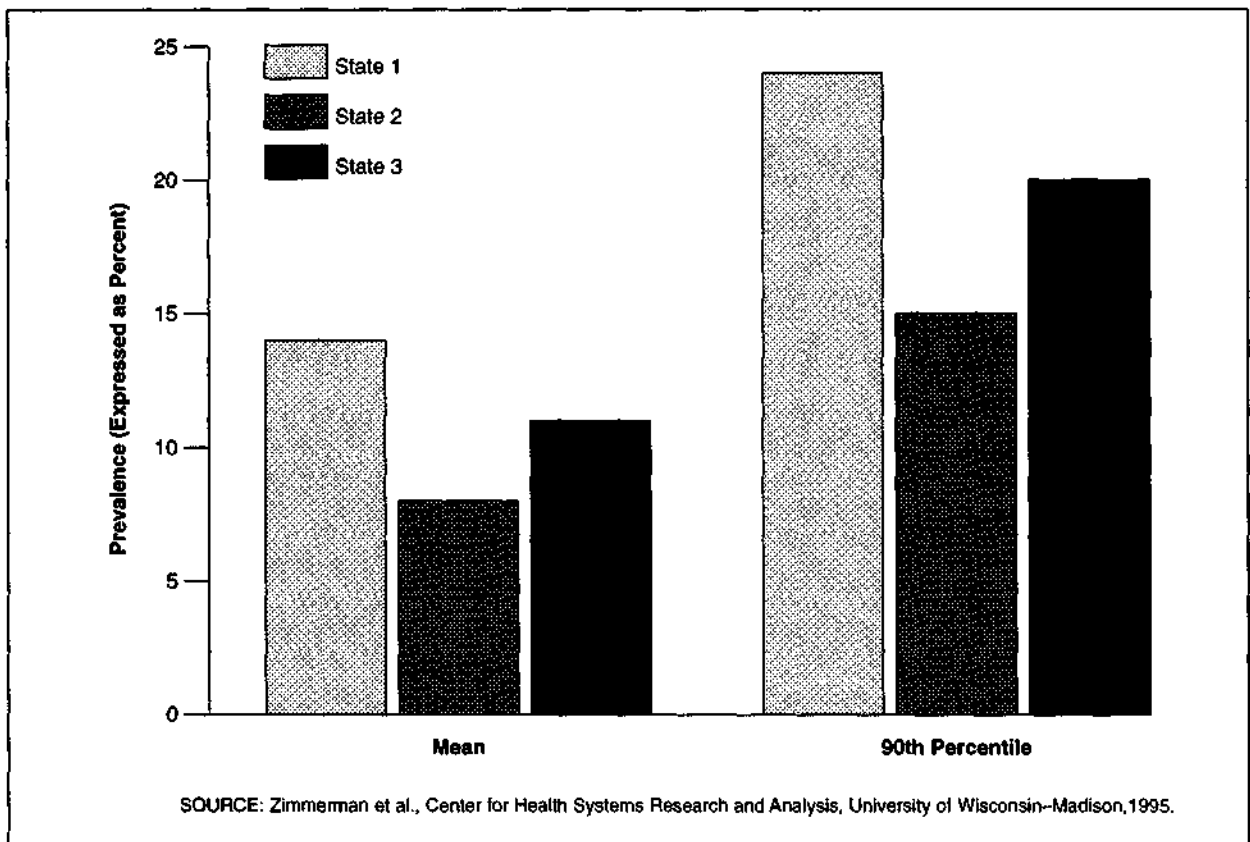


The determination of a threshold is closely related to the distribution of each QI at the facility level. Findings show that each QI distribution can be identified as belonging to one of three types—general indicators, rare events, or sentinel events—and that each type of distribution may imply a different method for determining threshold levels. General indicators are those QIs that have a fairly normal distribution, where thresholds can be easily set at a relative level. Rare events also may be normally distributed across facilities but within a smaller range and with more facilities likely to have no occurrence of the event. In these cases, thresholds may be set at some relative level as well or may be set at some absolute level by the number of occurrences, adjusted for facility size. Sentinel

events are not expected to occur, with most facilities having no occurrences of the QI. Regardless of the distribution, sentinel events are considered of such importance that a single occurrence in any size facility may be sufficient to suggest a potential quality-of-care problem.

In using relative thresholds, the choice of peer group is of key importance. It obviously is possible to define peer groups in many ways, e.g., based on ownership status, geographic region, facility size, hospital affiliation, or average facility case mix (in which case, risk adjustment might be unnecessary). Within the demonstration, we have considered two alternative peer groups: one that consists of all facilities in the four Medicare/Medicaid demonstration States, and another where each State forms its own

Figure 2
Prevalence of Falls Quality Indicator: Differences in Mean and 90th Percentile
in 3 Demonstration States



peer group. The selection of peer group can have a dramatic impact on the setting of a threshold and the consequent likelihood that a facility will be identified as having a potential quality problem related to any given QI. For instance, assuming a threshold set at the 90th percentile of facility QI scores in the peer group, in the latter half of 1993 the demonstrationwide threshold for the incidence of contractures was 25.0 percent; the State-specific thresholds for the same QI ranged from 19.0 to 37.5 percent. Another example is given in Figure 2 for the prevalence-of-falls QI. Figure 2 shows that the mean for State 1 (14 percent) on this QI is almost equal to the 90th percentile for State 2 (15 percent).

Although this variation in threshold levels may be the result of systematic State

differences in MDS+ accuracy, this is unlikely, because the definitions and training have been standardized across demonstration States. In general, the interstate differences in QI prevalence have declined over time, as facility staff have become more experienced in the use of MDS+ instruments. More likely, this variation is the result of care problems and priorities. State variation in QI prevalence points out a valuable use of the QIs as the basis for curriculum planning in cases where a State average indicates the possibility of widespread problems.

The use of alternative thresholds affects the number of QIs flagged in a particular facility, the survey resources required, and the comparative standing of different States. Again assuming a threshold set at the 90th

percentile in the latter half of 1993, if one were to use a demonstrationwide threshold, the average number of QIs flagged in a facility ranged across States from 0.9 to 4.0; using State-specific thresholds narrowed that range to 3.2-4.1. It also changed the ordering of the States, from lowest to highest average number of QIs for which a facility exceeded the threshold; i.e., the State that had the lowest average number of QI facility flags (0.9) when a demonstrationwide threshold was used was not the same State that had the lowest average (3.2 QIs) using the State-specific threshold.

Consistent with our policy of adopting a conservative approach in the early stage of the demonstration, pending further empirical analysis, we plan to use the State-specific 90th percentile as the threshold level for targeting a potentially problematic care area in a facility. The establishment of thresholds is the subject of continuing analysis and is being explored as part of the study of QI validity.

Target Efficiency

Another methodological concern addresses what we have called the "target efficiency" of the QI. This issue involves the specificity and sensitivity of the QI, in particular the likelihood of a false positive, i.e., that the QI will identify a resident or a facility for whom the QI flag is not ultimately found to represent a problem with the quality of care. Minimizing the number of false positives and false negatives is a critical concern, because each one decreases both the effectiveness and the efficiency of the quality-monitoring process. False positives also may promote an erroneous perception of a quality-of-care problem for a facility where no such problem exists. Using too strict a QI definition, however, may result in the opposite problem, failing to identify quality problems that, in fact, exist.

The target efficiency of the QIs varies with the extent to which the QIs: (1) are prevalence versus incidence measures; (2) include both process and outcome measures; and (3) can be risk adjusted. For instance, the QI indicating the prevalence of pressure ulcers does not possess target efficiency. On the other hand, a related-incidence QI adjusted for risk can be defined as the presence of risk factors for pressure ulcers at one point in time, followed by the development of pressure ulcers at the following point in time. Such a QI is highly efficient for targeting potential care problems in facilities. Similarly, when combined with a process measure, QIs can become more efficient. For example, the QI defined as "high risk of pressure ulcers and no skin care program" is more efficient than the simple prevalence of pressure ulcers. The most efficient QI for this particular care issue is defined as the high risk of pressure ulcers, with no skin care program, followed by the development of pressure ulcers. In general, even preliminary conclusions about the target efficiency of the QIs require validation studies, which are currently underway.

A second example of issues in target efficiency is provided by QIs related to the use of antipsychotic drug use. From lowest to highest target efficiency, these can be defined as:

- Antipsychotic drug use.
- Antipsychotic drug use for an extended period.
- Antipsychotic drug use for an extended period with cognitive decline, increased falls, or adverse medical outcomes.

An interesting distinction between these two examples is that the pressure ulcer illustration increases the efficiency of an outcome QI by adding process considerations, whereas in the antipsychotic drug case, target efficiency is increased by adding outcome considerations to a process QI.

In the development and selection of QIs for use in the demonstration, we have chosen to use fairly simple measures, rather than those that we believe have the greatest target efficiency. This decision is based on several considerations. First, the more target-efficient QIs are often difficult to interpret, because of their complex definitions. Second, use of more target-efficient QIs may result in an exclusion of cases that are a result of poor quality of care but that do not meet all of the conditions set forth in the complex QI definitions, thereby resulting in an increase in false negatives. Third, the use of complex definitions to increase target efficiency may also result in increased error. Specifically, any error that results from the first component of a complex definition can be multiplied as the remainder of the definition compounds the error. Finally, the use of the QIs in the monitoring process can take advantage of the survey as a source of immediate verification, detecting false positives. The important general point with respect to target efficiency is that the more likely the case that the indicator itself is to be used to render decisions on quality of care without followup or verification, the more important is the target efficiency of that indicator.

QI Validity

Given the potential impact and intended uses of the QIs, it is essential that their validity be established. The process of QI development, with its combination of empirical analysis and clinical input, has contributed to a high level of face validity. As part of the implementation of the quality component of the demonstration, we are conducting a more rigorous assessment of the accuracy and validity of the QIs. This validation study is designed to answer the following questions about the QIs:

- Are the data and algorithms used to construct the QIs accurate?
- Does the QI correctly indicate a problem with the quality of care at the resident level (i.e., for the specific resident in the case being investigated)?
- If there is a quality-of-care problem at the resident level, is it of sufficient severity and/or scope as to indicate a problem at the facility level that needs to be addressed?
- Are the problems identified, either on the basis of severity or scope, at a level that would warrant the citation of a deficiency under Federal regulations?

To address these questions, we are conducting a series of validation studies. At this point, 20 studies are planned across 5 of the 6 demonstration States. Each study involves a team of two validation team members. These individuals are respected experts in their fields, who have experience in the survey process, who are aware of the kinds of standards that typically would apply, and who have had extensive experience in the Federal survey process applying Federal standards in all aspects of care, including nursing, nutrition, medications, quality of life, and resident assessment. Validation team members spend an average of 3 days in a nursing facility. During that time, they assess a set of four to six QIs, preselected on the basis of facility QI reports that are prepared in advance. These QIs are assessed through a review of approximately 25 individual resident cases, using a combination of resident observation, resident and staff interviews, and record review. Most of the preselected QIs are ones on which the facility appears to be an outlier. In addition, each study includes consideration of a QI that does not appear to be an outlier. Thus, the design allows us to concentrate on the sensitivity (true or false positives) of the QIs but also permits an examination of their

specificity (true or false negatives). In addition, the information can provide insights on our choice of threshold levels. Each QI will be validated in at least three facility studies.

The validation team members review QI accuracy based on the MDS+ data by reviewing the relevant MDS+ instrument and by determining whether it is consistent with other information in the clinical records. Where possible, members also consider information from conversations with facility staff, residents, and family members. The determination of whether or not a QI represents a true quality-of-care problem is made by the validation team members, using information gathering and decision procedures similar to those used in the Federal survey process, but more specifically focused on the care area(s) covered by the QI. The team first determines whether a care problem exists at the individual resident level, and then whether the severity or scope of the problem is sufficient to conclude that a facility-level problem exists.

Validation studies are conducted concurrent with the regularly scheduled survey, although the validation team and survey teams do not interact except as is necessary for logistical purposes. By scheduling the validation studies in this manner, we are able to compare the validation teams' findings with those of the survey team. This provides a form of concurrent validity. It is important to note, however, that the survey findings are not assumed to be a "gold standard." For a variety of reasons, the survey results may not be consistent with the validation team findings, especially at the level of detail associated with determining quality in particular areas of care.

ISSUES FOR FURTHER RESEARCH

In addition to the current work on QIs and their validity, several directions for future work are being pursued. These

include setting thresholds, adjusting for risk, and aggregating the QIs into composite scores.

Continued Work on QI Thresholds

At this point, we are using a suggested threshold of performance equal to the 90th percentile within each State. The validation studies now being conducted will provide us with some information on the appropriateness of that threshold. Further study is needed to determine the most appropriate threshold for each QI. It is not necessary that a single approach to setting thresholds be used for each QI. Differences in QI distributions and seriousness must be balanced with cost considerations. Issues of equity must also be considered, particularly with respect to the definition of a peer group, if relative thresholds are to be used. In some cases, an absolute threshold might be more appropriate. These issues must be the subject of further analysis, as well as discussion among policymakers, clinicians, and consumers.

Continued Work on Risk Adjustment

Much more work needs to be done on the appropriate treatment of risk factors in the application of the QIs. An important basic issue is the most appropriate method for adjusting for risk. Additional research needs to be undertaken to compare the impact of simple approaches (such as the one we have employed in treating risk as a dichotomous variable at the resident level) to more complicated approaches (in which a risk scale is constructed). Although such an approach would increase the complexity of the resident review issues, it would maintain the interpretive ease of using discrete risk groups (as opposed to a single QI score,

statistically adjusted for risk). Future research could consider the appropriate divisions into multiple risk categories and the consequences of such an approach. Similarly, the efficacy of treating risk-adjusted QIs as separate measures, as opposed to constructing a single measure, need to be evaluated further.

Another important risk-related research issue is the treatment of risk factors that the facility has some control over to intervene to mediate the relationship between the risk factor and the onset of the condition covered by the QI. Although we have made what we believe are reasonable decisions about which potential risk factors have a quality component inherent in them, more research is needed on this issue. Specifically, further work should address the questions of which risk elements are amenable to intervention by facility staff and which represent factors that are independent of such interventions.

The risk groups currently in use are QI-specific—the definition of high and low risk varies by the QI. Future research may consider whether there is a single set of risk factors that could be used for all QIs. It might be possible to create groups of residents who are at greater or lesser risk of all outcomes and processes of care measured by the QIs. The relative risks for each QI would not need to be of equal magnitude or, for that matter, even of equal order. One group of residents might be at increased risk for one outcome but lower risk of another, relative to another group of residents. Such groups should be constructed to maximize the within-group homogeneity and the between-groups heterogeneity, with regard to each QI. The optimal number of resident risk groups can be statistically determined. However, it is important that such groups also be clinically reasonable.

QI Aggregation

Another important research issue is the relationships among the QIs and, as a result, their aggregation potential. As with any outcome-measurement initiative, one of the most important considerations is whether the individual items in the set of QIs can be combined or aggregated in meaningful ways to form a composite score or index that can be used for comprehensive assessment of nursing home quality of care.

There are at least two important qualifications to keep in mind in taking on this task. The first is that, although the QIs provide broad coverage of the major areas of nursing home care, there are areas in which their coverage is limited, because of the coverage limitations of the MDS+ as a source instrument. In particular, the QIs cannot address some dimensions of quality-of-life issues, such as resident rights to dignity and privacy, financial management, etc. It should be noted in passing, however, that in many cases the QIs do provide some insight into these issues because they are important derivative considerations to items directly covered by the QIs. For example, a resident with pressure sores or incontinence may have significant privacy or dignity issues that arise secondary to those problems. Second, even in areas covered by the QIs, one must proceed cautiously in aggregating across domains (or in some cases even within them), because the aspects of care addressed are independent.

There is much work to be done to test the relationships between the QIs before they can be aggregated to an overall measure of quality. We are currently involved in investigating the interrelationships of QIs within and across domains as well as over time.