Validation of Long-Term and Post-Acute Care Quality Indicators

CMS Contract No:
500-95-0062/
T.O. #4

Executive Summary

July 15, 2002

Prepared for
Yael Harris, Project Officer
Centers for Medicare and Medicaid Services
Office of Clinical Standards and Quality / S3-02-01
7500 Security Boulevard
Baltimore, Maryland 21244

Principal Authors
John N. Morris
Terry Moore
Rich Jones
Vincent Mor
Joseph Angelelli
Katherine Berg
Christine Hale
Shirley Morris
Katharine M. Murphy
Melissa Rennison

Contract Agencies:
Abt Associates Inc. (*Prime Contractor)*
55 Wheeler Street
Cambridge, Massachusetts 02138-1168

HRCA Research and Training Institute
1200 Center Street
Roslindale, Massachusetts 02131

Brown University
Center for Gerontology and Health Care Research
171 Meeting Street, Box G-B213
Providence, Rhode Island 02912

Key Staff:
John N. Morris
Principal Investigator, HRCA
Katharine M. Murphy
Co-Investigator, HRCA
Vincent Mor
Co-Investigator, Brown University
Katherine Berg
Co-Investigator, Brown University
Terry Moore
Project Manager, Abt Associates

CMS Project Officer:
Yael Harris

Executive Summary

Assessments of health care quality and the dissemination of resulting information about quality is becoming more widespread in the U.S. These assessments are most frequently in the form of "quality indicators" that are intended to reflect the quality of the care delivered or the patient care outcomes that can be attributed to the care delivered by various health care providers. In this report, we report on the results of our efforts to validate a series of quality indicators for use with chronic and post-acute care nursing home residents. Some of the other sources of quality indicators that are in current use include the Agency for Health Care Quality's Inpatient and Prevention Quality Indicators, the CAHPS (Consumer Assessment of Health Plans), the National Committee for Quality Assurance's HEDIS measures, Outcome-based Quality Indicators (OBQIs) for home health, dialysis care quality measures, and nursing facility quality indicators. The development of the latter three types of measures have been funded by the Centers for Medicare and Medicaid Services (CMS) in whole or in part, and federal initiatives are underway to utilize the home health and nursing facility measures for regulatory as well as public reporting purposes. Dialysis care measures are currently publicly reported on the CMS website.

The types of measurement information commonly utilized in making judgments about the value of a particular quality indicator include whether the measure has face (or clinical) validity and construct validity, and whether it reliably captures and measures what it purports to measure (validity). Earlier work under this contract (the CMS-sponsored "Development and Validation of Long-term and Post-acute Care Quality Indictors" project) established a set of 45 Minimum Data Set-based (MDS) quality indicators for use in nursing facilities that fulfilled select measurement criteria such as those cited above. These indicators were provisionally recommended for use by CMS, pending an assessment of their reliability and validity (Abt Associates Inc., 2001).

This report summarizes work performed to date to validate these 45 existing and newly developed quality indicators for the long-term and post-acute care populations residing in nursing facilities. Thirty indicators applicable to the chronic (or long-term) care population that were originally developed by others were evaluated, as were 15 newly developed measures for the chronic and post-acute care populations[1]. To our knowledge, the only previous work done to validate any nursing home quality indicators of this type was performed by the Centers for Health Research and Analysis at the University of Wisconsin[2] (see Zimmerman et al., 1995; Zimmerman et al., 1999; and Zimmerman and Karon, 1997). The list of measures examined in this study may be found in Tables 1 and 2.

Many of the indicators studied here are already in use by CMS in the quality monitoring system utilized in the long-term care survey process. Many facilities actively use these measures for

[1] Original developers of existing quality indicators examined in this report include the Centers for Health Research and Analysis, University of Wisconsin, LTCQ Inc., and J.D. Ramsey.

[2] "Validate" in this context means to clinically review the indicator against medical record and other primary data. Other developers may have performed other types of validation, for example, through secondary data analysis or the convening of industry experts, but this is not the type of validation done in this study nor in the Centers for Health Research and Analysis validation studies.

enhancing internal quality performance. In addition, nine of the measures reported upon here are being publicly reported for the states of Colorado, Florida, Maryland, Ohio, Rhode Island, and Washington as part of the CMS "nursing home quality initiative"[3].

Defining nursing home quality

Nursing facility quality is multidimensional, encompassing clinical, functional, psychosocial and other aspects of resident health and well-being. In this examination of nursing facility quality indicators, all listed aspects of resident functioning are addressed in varying degrees, and the needs of chronic residents and post-acute patients are separately examined. In most cases, multiple quality indicators (QIs) are recommended within a given domain of quality (e.g., clinical quality), and we propose that CMS utilize several QIs from each domain for purposes of public reporting, quality monitoring, and performance improvement. As stated previously, quality of care is necessarily multidimensional. No single QI is likely to capture overall facility quality. Facilities may perform extremely well on one type of QI, but may not perform nearly as well on another. Indeed, two papers recently confirmed this hypothesis, one using New York state data and the other data from Massachusetts (Mukamel and Brower, 1998; Porell and Caro, 1998). It is therefore important to present different indicators across multiple domains for a full view of facility quality performance.

Measurement of quality

The research design utilized in this quality indicator validation study follows that of other researchers who have concluded that quality must be measured by examining the interaction of structural, process and outcome measures (Donabedian, 1980; Sainfort et al, 1995; Zimmerman et al, 1995; Ramsey, et al, 1995). Each of these quality dimensions was incorporated into our hypotheses concerning the factors that enable a facility to prevent clinical and other problems from occurring, our subsequent collection of data from nursing facilities, and the analyses of these data.

Validation Study Parameters

The final analytic sample for this study was comprised of 209 freestanding and hospital-based facilities located in six states: California, Illinois, Missouri, Ohio, Pennsylvania and Tennessee. Facilities were selected for participation in the study based upon their quality indicator scores (observed over the prior year) on twenty QIs, their geographic location and their willingness to participate in the data collection protocols. The total patient sample included in our on-site field review comprised some 5,758 chronic and post-acute patients, although these facilities serve over 20,000 residents at any one point in time. Compared to all facilities in the states from which they were selected, participating facilities tended to be somewhat larger, were more likely to be non-profit and were less likely to be located in rural settings.

Both primary resident-level and facility-level data were collected in each sampled facility. At the resident level, medical records were reviewed to determine care processes provided to a representative resident sample during the time period of interest in twenty-one quality

[3] These are designated as "pilot" quality indicators and are listed in Tables 1 and 2.

dimensions, such as physical restraint use, pressure ulcers and pain. The types of care processes reviewed included whether comprehensive assessments other than the MDS were performed, whether physicians were notified in a timely manner following resident change in status, and whether care planning was documented in the record for identified problems. In addition, a subset of MDS items was independently assessed by research nurses for later comparison to facility MDS assessments. Facility-level data collected included an administrative survey in which questions were asked of the administrator and director of nursing, and an observation of the general facility environment.

Methods

Description of the Quality Indicators

In constructing the set of quality measures evaluated here, there has been a concern for possible inter-facility variation in the types of residents admitted and served by the facility; difference in the mix of residents served across facilities raises the possibility that inter-facility comparisons may be biased. To control for this possibility, where deemed necessary, three adjustment strategies have been applied.

1) For all of the indicators a denominator exclusion rule was applied (e.g., residents near death). These residents were not considered in the calculation of the quality indicator.
2) For four of the CHSRA indicators, two sub-versions of the same overall indicator were created for each facility, one applying to high-risk residents, the other to low risk residents. In addition, an overall high/low risk indicator was calculated.
3) For many of the other indicators, including those created by the project team and LTCQ Inc., some type of statistical regression-based covariate adjustment strategy was employed. For many indicators, this involved traditional resident-level covariates, supplemented in many instances by a new type of facility-based adjustment based upon resident characteristics upon admission. This is referred to as the "facility admission profile". QIs constructed using a facility admission profile are designated as such in Tables 1 and 2.

Testing the reliability of the Quality Indicators

In each participating facility research nurses sampled up to 30 residents records, observing and speaking with (if possible) the resident to complete a reduced form version of the MDS in order to allow for a comparison of the MDS based upon facility assessors and that completed by the research nurse assessor. The rationale for examining the reliability of the QI information across all our participating facilities was to allow for the possibility that poor data quality might compromise our ability to adequately test the validity of the QIs. Having information on the average reliability of the MDS data on which the QIs are based allowed us the possibility of excluding facilities with poor data quality from the analyses.

Over 100 MDS data elements were incorporated in the reliability study. A kappa statistic was used to calculate the level of agreement between the facility and our research nurse assessor. This statistic is more stringent than merely calculating the percentage agreement because it

adjusts for the possibility of chance agreement that can occur if the condition in question is relatively rare (something true for many of the QIs).

Validation of the Quality Indicators

During the development of data collection tools for this study, expert clinical panels were convened to develop empirically based hypotheses about what constitutes quality of care in a given dimension (e.g., pain, activities of daily living (ADL)). This effort met with varied success, as there appear to be relatively few well-studied, research-based "standards of care" in use in the nursing facility environment[4]. In cases in which no empirical evidence could drive theories about what components of care qualify a nursing facility as a "good" performer, the expert panels created their own hypotheses. These hypotheses were then utilized to 1) develop data collection instruments to assess nursing home care processes and structures, and 2) direct analysis of these data.

For the primary validation task, individual validation elements, as well as a series of summary scales, from the three data collection tools (Medical Record Review, Administrative Survey, and Environmental Observation) were categorized by quality of care construct (or hypothesis), and then evaluated to determine the degree of their relationship to each quality indicator. The final categorization of quality of care constructs were defined as "preventive" and "responsive".

- ***Preventive*** strategies represent the class of actions that "good" facilities choose to follow in an attempt to minimize the emergence of problems; these strategies are anticipatory in character. Data elements categorized into the preventive construct include staff training, higher staff resource levels, and facility efforts at continuous quality improvement (CQI).

- ***Responsive*** strategies represent actions that facilities are likely to use as they recognize that residents have ongoing or emerging problems in different quality areas. They represent a service response "audit trail," and as such confirm that staff have recognized the problem. Externally, these facilities will be observed to have higher QI scores, but the medical record will reflect a recognition that action must be taken in response to identified resident problems. Examples of data elements gathered on-site that are categorized as responsive are the documentation of comprehensive assessments (other than the MDS), documentation of changes in resident status, and referrals to specialists from inside and outside of the facility (e.g., physicians).

In summary, preventive strategies work to reduce the prevalence or incidence of quality problems measured by the QIs. On the other hand, responsive strategies reflect the fact that quality problems may have emerged in the resident population and as such reflect a "failure" of the facility to prevent the problem (or failure to achieve expected improvement outcomes). Consequently, responsive strategies are associated with an increased prevalence of problems (i.e. quality indicators).

[4] The best examples found of empirically-based nursing facility care practices came from clinical guidelines established by the Agency for Health Research and Quality, such as the Pain Clinical Practice Guidelines.

While the constructs created from the various sources of data were conceptualized as falling into one class or another, clinically and administratively relevant data elements thought to be related to particular QIs might have been able to be classified as either preventive or responsive. Thus, our final classification of the validation elements was done based both upon how they related to one another as well as how they related to the QIs. While seeming to represent a "circular" logic (i.e. using one construct to validate another and then to apply the same logic in the other direction), this is a process that characterizes most efforts at construct validation. Thus, the validation elements and constructs were examined for directionality relative to selected QIs and then the QIs were each formally tested against the battery of constructs (classified as preventive or responsive) to determine whether facilities records and care processes and structures related to the QIs in the expected direction.

For each of the constructs or individual data elements categorized as preventive or responsive, the relationship between it and the full array of quality indicators under study was reviewed. To be found acceptable, the construct had to have a consistent relationship across multiple quality indicators. For example, to be classified as preventive, a data element (e.g., a CQI monitoring protocol) had to be related to several quality indicators. We required that there be a clear directional relationship between the construct and the quality indictors. Specifically, preventive elements had to always show a positive relationship to lower (less problematic) QI rates, while responsive elements had to demonstrate a positive relationship to higher (more problematic) QI rates. In other words, the correlation between preventive elements and quality indicators had to be negative, and the correlation between responsive elements and quality indicator scores had to be positive to be considered clearly directional.

In evaluating the validity of the quality indicators, several summary measures were created:

1) a count of the number of significant preventive or responsive validation elements for the quality indicator, with the greater the count, the greater the confidence in the relationship;
2) a measure of the pooled association of the list of significant validation elements with the quality indicator. The latter is derived from a regression equation, and in this case represented by the multiple correlation coefficient. This is a multivariate-derived value that resembles a standard bivariate correlation[5]. In reviewing these values, we settled on a combination of two factors in assigning each of the candidate quality indicators to one of three "valid" categories: Top, Mid, and Not Validated; and
3) the underlying reliability of the MDS item and resulting QI.

To understand how these preventive and responsive factors were applied in establishing the validity of a QI, we provide examples of how these elements individually relate to two of the chronic quality indicators, "Pressure ulcer prevalence" (high & low risk) and "ADL worsening". Both indicators are assigned to the Level I, Top Validity category, and both achieved this status on the basis of the preventive elements alone. For the Pressure ulcer indicator, there was also a substantial array of individual responsive relationships, while for the other, ADL worsening, there was only one item of this type.

[5] Note: this value can be squared to get the classic R2 estimate of explained variance.

The ***Pressure ulcer*** quality indicator quantifies the proportion of at-risk residents in a facility that have a pressure ulcer (i.e., bed sore, decubitus ulcer, pressure sore) of severity ranging from one persistent area of redness that does not disappear when pressure is relieved to one or more open wounds where the full thickness of skin and subcutaneous tissue is lost and underlying bone or muscle is exposed.

There are a large number of clinical and functional risk factors for pressure ulcers (e.g., poor nutrition, incontinence, diabetes, immobility); thus, a number of preventive activities and responsive factors were evaluated. Preventive activities, in general, relate to the handling of at-risk residents and treatment of conditions that contribute to or mitigate pressure ulcer risk. Responsive activities, in general, define actions that a facility's caregivers take to document, communicate and attempt to ameliorate pressure ulcers once present.

Preventive activities for pressure ulcer prevalence included the screening, assessment, and treatment for conditions placing residents at risk for pressure ulcers. Thus, the following individual data elements or constructs were found to be associated with lower pressure ulcer prevalence:

- More frequent scheduling of assessments for suspicious skin areas.
- Weekly routine assessment using a standard protocol for delirium, that would - if present - keep residents bed-bound.
- Observations on the environmental assessment of residents walking or otherwise out of bed.
- Observations on the environmental assessment of caregivers providing assistance to residents with nutritional needs.
- A constructed scale expressing the extent to which a facility manages clinical, psychosocial, and nutritional complications across domains in a manner consistent with high quality care (expressed as a single factor score).

Staffing factors provide additional (albeit indirect) evidence of preventive activities. For example, staffing items related to pressure ulcer prevalence were 1) the absence of facility management change; and 2) the extent that a facility did *not* rely upon floats or contract staff.

Responsive activities for pressure ulcer prevalence include policies, procedures or actions taken by caregivers in response to existing or newly detected pressure ulcers. Identified activities include:

- Comprehensive assessment (other than the MDS) of pressure ulcers documented in the medical record.
- Assessment of pressure ulcers by a physician.
- Clear documentation in the medical record that the resident has a problem in this area or that the resident's condition has changed relative to pressure ulcers.
- Where change was noted in the medical record, there is documentation that this change 1) was evaluated within 72 hours, 2) resulted in a notification to physician or therapist, 3) resulted in a referral to a consultant, and/or 4) resulted in a change in the care plan.

An additional theme related to pressure ulcers was a constructed measure of the extent to which the medical record and care plan agree that pressure ulcers are a problem. This level of agreement signals facilities with a well-integrated system for problem recognition and treatment implementation.

For ***ADL worsening***, there were 17 significant preventive elements and one significant responsive element. From this set of preventive elements, three primary themes emerge: attention to the resident as an individual, an engaging and safe environment, and good continence care. Further explanation of these themes and related data elements follows.

- Maintaining ADL gains is related to a concern with what the resident is thinking and who he or she may be as a person, as seen in areas related to cognition, behavior, and pain. Better outcomes (i.e. facilities have lower rates of ADL worsening) are observed when there are: 1) CQI monitoring protocols in place for behavioral function and communication; 2) weekly routine screening of communication and pain, using standard protocols; and 3) rooms that are personalized with furniture, photos, and other things from the resident's past.

- Maintenance of ADLs is also related to an environment in which the resident is up and out of bed and engaged in activities. Better outcomes are related to a series of things that were observed by the research nurse about the facility, including: 1) residents being up and about; 2) residents seen to be walking or independently moving about the facility with or without assistive devices; and 3) indications that a variety of activities are available for residents with different capabilities. Related data elements observed during inspection of the facility environment were that public and common areas were well lighted and resident safety had been considered.

- Finally, there was a link to facility efforts aimed at good continence care. Preventive elements relating to this theme include: 1) a scale that counted up to 15 "good" incontinence management items; 2) a scale that focused on care practices relevant to promoting improved levels of continence; 3) a scale that looked specifically at ADL training approaches that were targeted to helping residents maintain continence patterns; and 4) a CQI monitoring protocol in place for bladder incontinence.

Findings

Reliability was evaluated in several ways. Research nurse MDS assessments were compared to facility-generated MDS assessments to generate the following statistics: 1) percent agreement between "gold" standard nurses and facility nurses; 2) MDS item-level kappas; and 3) kappas for a subset of the QI where these could be established (i.e., for prevalence QIs only).

Table 1 displays reliability and distributional statistics for each of the quality indicators for the 209 facilities in the national study sample. Reliability was assessed using the weighted kappa statistic, with a value of .40 or higher being considered indicative of inter-assessor agreement, while a value of .75 or higher is indicative of superior inter-assessor reliability. In this case the weighted kappas reflect the cross-sectional reliability of the MDS items that comprise the numerator of the quality indicator (e.g., the numerator for the "Falls" QI is MDS item J4a). Using this standard, only one of the MDS items for a QI numerator falls below the .40 threshold (MDS item N2, which makes up the "Little to no activity" QI). Thirty-one of the quality indicators are based on MDS items with an average weighted kappa of .70 or higher.

Table 1 also displays the mean rates of the quality indicators across the 209 sampled facilities. As seen here, only two quality indicators have very low prevalence (i.e. < 5%). The rate of the chronic care "New insertion of indwelling catheter" indicator is two percent, and the rate of the post-acute care "Failure to improve and manage delirium" indicator is three percent across the sampled facilities. Other QI rates range between five and 92 percent. The rate of occurrence of various QIs is another criterion that should be taken into consideration when evaluating the utility of various QIs.

An exploration of the presence of "measurement bias" was also completed, in order to understand whether particular QIs are more subject to over- or under-reporting by facilities than others. If this were the case, we would be able to evaluate the ability of a facility-level adjustment mechanism (referred to here as a "facility admission profile") to capture this measurement bias. Methods and findings for this aspect of the validation study may be found in the body of this report. By and large, while there was inter-state variation in the extent of over or under-reporting, relatively few QIs were observed to have large levels of under or over-reporting in general and relatively few facilities were systematically over or under-reporting the prevalence of quality problems as measured by a multiplicity of QIs.

Table 2 displays the summary measures of quality indicator validity. Of the master list of 45 quality indicators, 2 could not be evaluated due to missing quality indicator data.[6] Thirteen of the chronic care indicators and four of the post-acute care indicators were judged to be in the Level I (Top) validation category. An additional group of sixteen chronic and two post-acute indicators were also accepted as valid, and placed into Level II, the Mid-Valid Category. A total of seven chronic care indicators and one post-acute care indicator were judged not to be valid.

Conclusions and Recommendations

In this national validation study, there is strong evidence that many of the set of 45 reviewed quality indicators capture meaningful aspects of nursing facility performance. We highly recommend for use by CMS and nursing facilities any of the QIs that fall into the Level I validation category, as these QIs have the strongest degree of evidence that they represent real care processes in nursing facilities. The chronic care quality indicators with the highest level of validity include:

- Prevalence of indwelling catheter;
- Bladder/bowel incontinence (high and low risk, high risk, low risk);
- Urinary tract infections;
- Infections;
- Inadequate pain management;
- Pressure ulcers (high and low risk);
- Late-loss ADL worsening;
- ADL worsening;
- Locomotion worsening;

[6] High and low risk pressure ulcers will be evaluated separately and findings submitted upon delivery of the final validation report.

- Improvement in walking; and
- Worsening bladder continence.

Four post-acute care quality indicators are highly valid, including:

- Failure to improve and manage delirium[7];
- Inadequate pain management;
- Failure to improve during early post-acute period; and
- Improvement in walking.

The chronic quality indicators that we recommend rejecting for further use at this time are:

- Behavior symptoms (high risk and low risk);
- Weight loss;
- Antipsychotic use (high risk and low risk);
- Worsening behavior; and
- Worsening pressure ulcers.

The post-acute care indicator that proved not to be valid is "Failure to Prevent or Improve Pressure Ulcers" and therefore should be rejected for use by CMS.

Those QIs that fall into the Level II – Mid Valid category are deemed appropriate for use in measuring nursing facility quality, as they do offer evidence of validity; they are simply not as highly recommended to CMS as those QIs falling into the "Top" (Level I) validation category. In making final determinations about the utility of these QIs for performance improvement, public reporting or other purposes, CMS may want to review both the prevalence and the reliability of these indicators.

A special note is warranted on the "Little or No Activity" quality indicator. While based on the validation effort it was judged to fall into the Mid-Valid (Level II) category, the MDS item on which the indicator is based was found to have poor reliability. Should CMS choose to utilize this indicator for public reporting, facilities will need instruction on proper coding of this assessment item.

In addition to determining which of these sets of nursing facility quality indicators are "valid", or reflecting the care outcomes and issues they are purported to reflect, these results provide evidence that quality indicators measure aspects of care quality that may be amenable to modification through facility practice. For example, facility staffing and policies, practices or procedures are found to be related to resident quality outcomes and therefore may be modified by facilities to enhance quality of care delivery.

[7] Again, this QI has a very low rate of occurrence (three percent) in our study sample. The national distribution of this indicator should be examined as CMS makes a final determination as to this QI's overall utility.

Table 1

QI Rates and Weighted Kappas

Quality Indicator	QI Proportional Rate – The Average Across Facilities	Standard Deviation of the QI Rate	The Rate in the Facility with the Lowest Proportional Problem	The Rate in the Facility with the Highest Proportional Problem	Average Weighted Kappa for MDS Items Composing the QI [1]
Chronic Prevalence					
++Behavior symptoms (high&low risk)BEH1	.20	.10	.00	.68	.71
++Behavior symptoms (high risk) BEH2	.23	.11	.00	.69	.71
++Behavior symptoms (low risk) BEH3	.07	.05	.00	.23	.71
Little or no activity SOC2	.12	.12	.00	.77	.28
Prevalence of indwelling catheter CAT2	.07	.05	.00	.32	.71
++Bladder/bowel incontinence (high&low risk) CNT1	.62	.13	.14	.89	.88
++Bladder/bowel incontinence (high risk) CNT5	.93	.05	.76	.99	.88
++Bladder/bowel incontinence (low risk) CNT6	.49	.13	.12	.83	.88
Urinary tract infections CNT4	.08	.05	.00	.31	.53
Falls FAL1	.08	.04	.00	.24	.52
++Infections (pilot) INFX	.17	.08	.00	.43	.50
++Feeding Tubes NUT1	.08	.05	.00	.27	.80
++Low Body Mass Index BMIX	.12	.05	.00	.31	.85
++Weight loss (pilot) WGT1	.08	.04	.00	.26	.42
++Inadequate Pain Management	.11	.08	.00	.48	.73

Table 1

QI Rates and Weighted Kappas

Quality Indicator	QI Proportional Rate – The Average Across Facilities	Standard Deviation of the QI Rate	The Rate in the Facility with the Lowest Proportional Problem	The Rate in the Facility with the Highest Proportional Problem	Average Weighted Kappa for MDS Items Composing the QI [1]
(pilot) PAIX					
++Pressure ulcers (high&low risk) (pilot) PRU1	.09	.05	.00	.27	.74
++Pressure ulcers (high risk) PRU2	*	*	*	*	*
++Pressure ulcers (low risk) PRU3	*	*	*	*	*
++Burns, skin tears or cuts BURX	.05	.04	.00	.19	.46
Restraints used daily (pilot) RES1	.07	.09	.00	.49	.56
++Antipsychotic use (high&low risk) (pilot) DRG1	.21	.08	.02	.43	.89
++Antipsychotic use (high risk) DRG2	.43	.11	.26	.61	.89
++Antipsychotic use (low risk) DRG3	.17	.07	.02	.40	.89
Chronic Incidence					
Late-loss ADL worsening (pilot) ADL1	.16	.09	.00	.44	.84
ADL worsening ADL2	.08	.07	.00	.33	.83
ADL improvement ADL3	.25	.09	.08	.48	.83
++Locomotion worsening MOB1	.14	.07	.01	.40	.82
++Improvement in walking WALX	.82	.08	.61	.99	.84
++Cognition worsening COG1	.12	.07	.00	.43	.76
++Worsening communication	.11	.07	.00	.31	.83

Table 1

QI Rates and Weighted Kappas

Quality Indicator	QI Proportional Rate – The Average Across Facilities	Standard Deviation of the QI Rate	The Rate in the Facility with the Lowest Proportional Problem	The Rate in the Facility with the Highest Proportional Problem	Average Weighted Kappa for MDS Items Composing the QI [1]
COM1					
++Delirium DELX	.09	.06	.00	.29	.61
++Worsening behavior BEH4	.07	.05	.00	.24	.72
++Depressed anxious mood worsening MOD3	.15	.07	.00	.37	.60
New insertion of indwelling catheter CAT1	.02	.02	.00	.09	.71
Worsening bowel continence CNT2	.19	.09	.00	.41	.88
++Worsening bladder continence CNT3	.19	.09	.00	.49	.87
++Pain worsening PAN1	.10	.05	.00	.26	.73
++Worsening pressure ulcers PRU4	.07	.04	.00	.27	.74
Post-acute Prevalence					
++Failure to improve and manage delirium (pilot) DELX	.03	.03	.00	.16	.65
++Inadequate pain management (pilot) PAIX	.27	.10	.02	.60	.72
Post-acute Incidence					
Failure to improve during early post-acute period ADLX	.63	.19	.14	.99	.72
++Failure to improve bladder incontinence CNTX	.55	.09	.32	.79	.73

Table 1

QI Rates and Weighted Kappas

Quality Indicator	QI Proportional Rate – The Average Across Facilities	Standard Deviation of the QI Rate	The Rate in the Facility with the Lowest Proportional Problem	The Rate in the Facility with the Highest Proportional Problem	Average Weighted Kappa for MDS Items Composing the QI [1]
++Failure to prevent or improve pressure ulcers PRUX	.23	.09	.04	.50	.74
++Failure to improve or prevent respiratory problems RSPX	.92	.05	.77	.99	.53
++Improvement in Walking (pilot) WALX	.28	.14	.03	.71	.77

Notes:

1 Kappas below 0.4 reflect poor inter-rater reliability; a value between .40 and .60 is indicative of acceptable inter-assessor agreement; and a value of .75 or higher is indicative of superior inter-assessor reliability.

++ Quality indicator was risk-adjusted using facility admission profile.

* Validation analyses were not complete for these QIs.

Table 2

Summary Measures of Quality Indicator Validity

Quality Indicator	Count of Significant Preventive Data Elements[1]	Count of Significant Responsive/ Reactive Data Elements	Total Count of Significant Data Elements	Multiple R (Measure of Association) For Preventive Elements	Multiple R For Responsive Elements	Multiple R for All Elements	Degree of Validity[2] I TOP II MID III NOT Valid
Chronic Prevalence							
++Behavior symptoms (high&low risk) BEH1	3	4	7	.34	.31	.43	II
++Behavior symptoms (high risk) BEH2	1	3	4	.25	.30	.39	III
++Behavior symptoms (low risk) BEH3	0	0	0	--	--	--	III
Little or no activity SOC2	8	1	9	.39	.13	.44	II
Prevalence of indwelling catheter CAT2	5	6	11	.45	.71	.78	I
++Bladder/bowel incontinence (high&low risk) CNT1	7	3	10	.50	.45	.66	I
++Bladder/bowel incontinence (high risk) CNT5	8	2	10	.57	.35	.65	I
++Bladder/bowel incontinence (low risk) CNT6	5	3	8	.47	.31	.56	I
Urinary tract infections CNT4	7	8	15	.51	.41	.59	I
Falls FAL1	4	7	11	.27	.40	.50	II
++Infections (pilot) INFX	6	9	15	.46	.36	.53	I
++Feeding Tubes NUT1	7	8	15	.44	.40	.54	II
++Low Body Mass Index BMIX	6	1	7	.39	.20	.41	II
++Weight loss (pilot) WGT1	3	0	3	.27	--	.27	III

Table 2

Summary Measures of Quality Indicator Validity

Quality Indicator	Count of Significant Preventive Data Elements[1]	Count of Significant Responsive/ Reactive Data Elements	Total Count of Significant Data Elements	Multiple R (Measure of Association) For Preventive Elements	Multiple R For Responsive Elements	Multiple R for All Elements	Degree of Validity[2] I TOP II MID III NOT Valid
++Inadequate Pain Management (pilot) PAIX	5	4	9	.32	.67	.74	I
++Pressure ulcers (high&low risk) (pilot) PRU1	10	12	22	.48	.43	.59	I
++Pressure ulcers (high risk) PRU2	*	*	*	*	*	*	*
++Pressure ulcers (low risk) PRU3	*	*	*	*	*	*	*
++Burns, skin tears or cuts BURX	4	7	11	.30	.34	.47	II
Restraints used daily (pilot) RES1	3	7	10	.33	.48	.52	II
++Antipsychotic use (high&low risk) (pilot) DRG1	5	3	8	.32	.31	.47	II
++Antipsychotic use (high risk) DRG2	0	1	1	--	.31	.31	III
++Antipsychotic use (low risk) DRG3	1	3	4	.15	.35	.38	III
Chronic Incidence							
Late-loss ADL worsening (pilot) ADL1	13	1	14	.49	.26	.51	I
ADL worsening ADL2	17	1	18	.57	.07	.57	I
ADL improvement ADL3	5	0	5	.39	--	.39	II

Table 2

Summary Measures of Quality Indicator Validity

Quality Indicator	Count of Significant Preventive Data Elements[1]	Count of Significant Responsive/ Reactive Data Elements	Total Count of Significant Data Elements	Multiple R (Measure of Association) For Preventive Elements	Multiple R For Responsive Elements	Multiple R for All Elements	Degree of Validity[2] I TOP II MID III NOT Valid
++Locomotion worsening MOB1	8	1	9	.62	.09	.62	I
++Improvement in walking WALX	9	0	9	.64	--	.64	I
++Cognition worsening COG1	12	8	20	.40	.34	.52	II
++Worsening communication COM1	3	5	8	.29	.31	.41	II
++Delirium DELX	10	0	10	.40	--	.40	II
++Worsening behavior BEH4	1	1	2	.15	.17	.24	III
++Depressed anxious mood worsening MOD3	7	0	7	.31	--	.31	II
New insertion of indwelling catheter CAT1	8	6	14	.40	.24	.44	II
Worsening bowel continence CNT2	3	1	4	.25	.30	.45	II
++Worsening bladder continence CNT3	6	5	11	.39	.40	.63	I
++Pain worsening PAN1	10	5	15	.37	.40	.51	II
++Worsening pressure ulcers PRU4	3	2	5	.27	.23	.35	III
Post-acute Prevalence [3]							
++Failure to improve and manage delirium	6	3	9	.58	.36	.62	I

Table 2

Summary Measures of Quality Indicator Validity

Quality Indicator	Count of Significant Preventive Data Elements[1]	Count of Significant Responsive/ Reactive Data Elements	Total Count of Significant Data Elements	Multiple R (Measure of Association) For Preventive Elements	Multiple R For Responsive Elements	Multiple R for All Elements	Degree of Validity[2] I TOP II MID III NOT Valid
(pilot) DELX							
++Inadequate pain management (pilot) PAIX	5	2	7	.52	.36	.64	I
Post-acute Incidence							
Failure to improve during early post-acute period ADLX	9	0	9	.59	--	.59	I
++Failure to improve bladder incontinence CNTX	3	0	3	.37	--	.37	II
++Failure to prevent or improve pressure ulcers PRUX	1	0	1	.12	--	.12	III
++Failure to improve or prevent respiratory problems RSPX	2	0	2	.42	--	.42	II
++Improvement in Walking (pilot) WALX	4	0	4	.48	--	.48	I

Table 2

Summary Measures of Quality Indicator Validity

Quality Indicator	Count of Significant Preventive Data Elements[1]	Count of Significant Responsive/ Reactive Data Elements	Total Count of Significant Data Elements	Multiple R (Measure of Association) For Preventive Elements	Multiple R For Responsive Elements	Multiple R for All Elements	Degree of Validity[2] I TOP II MID III NOT Valid

Notes:

[1] An alpha significance level for the correlation between the validation element and the quality indicator of .09 or lower.

[2] Level I -- Preventive Multiple R Equal to or Greater than .45 – OR -- Total Multiple R equal to or greater than .55

Level II -- Preventive Multiple R Equal to or Greater than .30 – OR -- Total Multiple R equal to or greater than .40

Level III -- Preventive Multiple R Less than .30 – OR -- Total Multiple R less than .40

[3] The sample utilized in evaluation of the post-acute care QIs includes hospital-based transitional care units (TCUs) only [maximum N = 52 facilities]. At the same time, we note that this was one of two analytic samples that could have been used to evaluate the post-acute indicators. Under a second sampling strategy, the TCU sample could be supplemented through the addition of 104 chronic nursing facilities. In each of these facilities there were sufficient numbers of Medicare residents on which to calculate the post-acute quality indicators. Had this second sample approach been the primary strategy to be followed, rather than the TCU approach on which this task rests, the Failure to Prevent or Improve Pressure Ulcer quality indicator would not have been rejected. In fact it would have been placed in Level I, the highest validation category. At the other extreme, had this alternative approach been used, the Improvement in Walking quality indicator would have been placed in Level III, Not Validated.

++ Quality indicator was risk-adjusted using facility admission profile.

* Validation analyses were not complete for these QIs.