Abt Associates Inc.

Cambridge, MA
Lexington, MA
Hadley, MA
Bethesda, MD
Chicago, IL

Design and Validation of Post-Acute Care Quality Measures

Abt Associates Subcontract: RAND Prime Contract No. 500-00-0026 (TO 2)

Final Report

January 31, 2005

Prepared for
Debra Saliba, PhD
The RAND Corporation
1700 Main Street. M-28
Santa Monica, CA

Prepared by
Terry Moore
Ning Wu
David Kidder
Barbara Bell
David Warner
Louise Hadden
Yevgenia Mackiernan
John Morris
Richard Jones

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Internal Review

Project Director

Technical Reviewer

Management Reviewer

Contents

Executive Summary

1.0 Introduction

In May 2004, the RAND Corporation subcontracted with Abt Associates Inc. to address several tasks related to work done on the nursing home project "Development and Validation of Indicators and Measures of Quality and Appropriateness of Services Rendered in Post-acute and Long-term Care Settings" (the MegaQI Project). The scope of work for this subcontract included the following:

1. Refresh CMS' publicly reported nursing home quality measure (QM) data by providing national facility and resident-level QM files in June and September 2004.

2. Document and deliver to CMS the "Validation Database", developed by the MegaQI Team.

3. Review, analyze and recommend post-acute care (PAC) quality of care measures from existing chronic and PAC measures. This task included:

    - Exploring the feasibility of expanding the number of valid PAC QMs, by reviewing and testing the PAC measures originally considered by the MegaQI Team, a set of PAC QMs proposed by HRCA in a memo to CMS in February 2004, and certain chronic QMs that "have face validity for post acute residents."

    - For some of the PAC QMs, applying risk adjustment, as appropriate.

    - Testing the validity of the expanded list of PAC QMs on two components of the 209 nursing homes: a) transitional care units (TCUs), and b) a larger sample that would include TCUs and other nursing facilities.

2.0 Purpose and Overview

CMS and others have expressed concern about the dearth of valid PAC QMs. Therefore, CMS' principal objective for this project was to learn whether or not it would be feasible to expand the set of valid PAC QMs, using data and methods applied in the MegaQI (June 10,2003) Validation Report. In this report, the MegaQI Team recommended six PAC QMs. Subsequently, the National Quality Forum (NQF) conducted a review of the MegaQI team's proposed QMs. The NQF recommended four PAC QMs that addressed rehospitalization, inadequate pain management, delirium and walking improvement, though only three were subsequently publicly reported.[1] NQF also recommended against using the Facility Admission Profile (FAP) adjustment applied to some of the publicly reported PAC measures. PAC QMs that are now reported are:

- Percent of short-stay residents with delirium;
- Percent of short-stay residents who had moderate to severe pain; and
- Percent of short-stay residents whose pressure sores have not gotten better.

To support additional validation work, CMS asked Abt Associates to fully document the architecture of the Validation Database, the code used to create validation scales and elements, the specifications

[1] The rehospitalization measure developed by the MegaQI Team has not been publicly reported, because it was not adequately developed, constructed or tested.

and code for all proposed PAC QMs, procedures for selecting facilities for validation analyses, and statistical routines used to validate the proposed QMs. CMS also asked Abt to replicate Appendix Table M of the 2003 report "Validation of Long-Term and Post-Acute Care Quality Indicators" (the Validation Report).

3.0 Description of Validation Database and Work Completed to Submit as Deliverable

Description of the Validation Database

The objective of the MegaQI validation study was to demonstrate a correlation between measures of nursing facility processes of care and resident-based outcome measures of presumed quality of care (QMs). Facility-level structure and process of care measures were collected directly by trained research staff using structured interview protocols, medical record reviews, and a questionnaire administered to key facility informants (administrators, nursing directors). The final analytic sample consisted of 209 freestanding and hospital-based facilities located in six states: California, Illinois, Missouri, Ohio, Pennsylvania and Tennessee. Medical record review data were based on a sample of residents within each facility, with a goal of reviewing 30 residents' charts per facility. The total resident sample included in the on-site field review comprised some 5,758 residents.

Data collection occurred at both the facility and resident levels. Facility-level primary data were linked to QMs developed from MDS data with a crosswalk file of facility ID variables. For the resident-level file, unique records (and residents) are identified by state and facility ID.

Files in the Validation Database include:

- All the data collected at the resident level, both MDS reliability data and data from a structured **Medical Record Review** (**MRR**). (resident-level file)

- All of the data collected at the facility-level, from the **Facility Walk-Through** (on-site observation of staff and residents, to gain an overall understanding regarding whether the facility was "resident-centered", what the "feel" of the facility was, and what the nature of staff interactions with residents were) and from the **Administrative Survey (**questions regarding staffing, residents, specialists, clinical communication protocols, care planning and training and orientation of staff). (facility-level file)

- Identifiers for matching validation data to quarterly QMs (facility-level file)

- Scales developed by the MegaQI Validation Study Steering Committee to predict nursing home QM scores. Each scale is defined in Appendix F of the 2003 Validation Report. (facility-level file)

- Summary data aggregated from the medical record review (MRR), for all facilities (facility-level file)

- All of the variables that are available in the Validation Database for use in attempting to validate a QM (facility-level file)

Acquiring, Reviewing and Preparing the Validation Database

Abt programmers tried to replicate findings reported in Appendix Table M of the Validation Report, as well as Table 2 from the same report. [2] After sustained effort and research into the database versions and statistical methods used in the earlier analyses, Abt programmers were able to replicate the statistics in these two tables. On August 18, 2004, Abt submitted to CMS a CD containing the full Validation Database, together with documentation.

4.0 Methods

Selection of PAC QMs for Validation

The selection of PAC quality measures for the current project began from an initial list of 29 candidate PAC QMs. These included publicly reported measures, measures developed and tested for validation by the MegaQI project but not publicly reported, new measures proposed by HRCA in a letter to CMS in January 2004 (see Appendix 6) and chronic care QMs that the team considered relevant to post-acute care.

Once the initial set of 29 measures was decided upon, the technical specifications were drafted, SAS code written, and the measures were processed against the MDS repository data for the period 2001-2002. We then examined the distribution of the PAC QMs and found three to fail our criteria for selection (insufficient or excessive prevalence). The three QMs dropped from further analysis were:

- COM02—No decline in communicative function
- RES01—Restraints (physical) used daily, prevalence
- COG01—Cognition worsening

Processing the Recommended QMs

We computed the recommended QMs on all MDS records for the US for four target quarters (2001Q3, 2001Q4, 2002Q1 and 2002Q2). Each QM could have at most two versions: adjusted (using regression techniques, if resident-level covariates were recommended) and unadjusted (except for exclusions from the numerator or denominator).

An "unadjusted QM" is calculated as the total number of residents for whom the QM is triggered, divided by the total number of residents who are at risk For PAC QMs that measure delirium, ADL, balance, mobility, incontinence and walking, we excluded three groups of residents from the denominators and numerators: residents in a coma, residents with end stage disease, and residents enrolled in hospice.

For some QMs, we adjusted for risk using resident-level covariates that have been found to increase the risk for an outcome. An "adjusted QM" compares the "observed" (unadjusted) QM to an "expected" QM, based on a resident's covariate values and the estimated coefficients from a prediction model. Covariates included items from the MDS, as well as indices and scales derived from the Resource Utilization Group-III (RUG-III) system.

Then we merged a file of unadjusted and adjusted QMs to data from the Validation Database for all facilities in the Validation Sample. In order to link chronologically each facility's QM measure to the

[2] Table 2 presented validation findings, including estimated levels of validity and the statistics on which these levels were based.

validation items from the Validation Database, we chose the date of the facility walk-through as a reference date. As a check on our work, we examined the QM distributions, and found no cause for concern. The distributions of the three currently publicly reported PAC QMs (DEL0X1, PAI0X1 and PRU0X2) were similar to the values presently publicly reported. Families of QMs representing the same clinical or functional areas (e.g., pressure ulcers) had similar distributions regardless of their respective risk-adjustment models. We also compared the QMs for all nursing facilities in the six sampled validation states (from which the sample of 209 were selected).

PAC Facility Selection

With the expectation that expanding the facility sample might increase the number of PAC QMs with statistically significant relationships to validation measures, CMS asked the project team to validate proposed PAC QMs using data from two facility samples:

- Transitional care units (TCUs, originally used to validate PAC QMs in 2002); and
- The full sample of nursing facilities. For PAC analyses, the full sample is defined as 196 facilities (facilities with one or more 5-day assessments in the data collection period for this study, from July 1, 2001 through June 30, 2002).

Adding more facilities could also create measurement problems. PAC QMs are meant to capture outcomes of care provided to short-stay residents who have potential for stabilization or rehabilitation. Yet there may be some facilities that do not provide "PAC-like" care to their short-stay residents. (For example, some might serve as way stations for residents who receive little rehabilitation, but move on to other settings after relatively brief stays). In the first Validation Report, TCUs were chosen to validate PAC QMs because it was presumed that their affiliation with hospitals meant that they would, for the most part, specialize in post-acute rehabilitative care. If the sample was expanded beyond TCUs, were there ways to infer from the available data which facilities were more likely than others to be providing PAC care to their short-stay residents?

The team and CMS agreed on some "behavioral" dimensions of short-term nursing facility care that could be used to refine the sample. On the assumption that PAC care plans should aim to stabilize or rehabilitate, facilities that adhere to PAC-like care practices should not admit large numbers of short-stay residents with very little or no potential for stabilization or rehabilitation. The team explored distributions of key variables and set thresholds to define "high percentages" (the highest 10 percent of the distribution). A facility may be said to adhere to a PAC model if it does not admit high percentages of short-stay residents:

- with severe cognitive impairment; (facility percent of 5-day assessments with RUG Cognitive Performance Scale (CPS) score >= 4) > 0.40
- with severe late-loss ADL limitations; (facility percent of 5-day assessments with RUG Late-loss Activities of Daily Living (ADL) score >= 16) > 0.45
- with high levels of clinical complexity; (facility percent of 5-day assessments with RUG Clinical Complexity Score (CCS) = 3) > 0.26
- in hospice care (facility percent of 5-day assessments with P1ao checked) > 0.01, who are end stage (facility percent of 5-day assessments with J5c checked) > 0.07, or are comatose (facility percent of 5-day assessments with B1 = 1) > 0.01.

Using these thresholds, the team decided to exclude any facility that shows up in three or more outlier categories and any facility with fewer than 20 14-day assessments. With these exclusions, the

"refined full" sample for the validation analyses included 182 facilities, including 56 TCU facilities (down from 60) and 126 non-TCU facilities (down from 136).

Validation Analyses

The primary goal of the validation analyses was to identify the PAC QMs that reflected the quality of post-acute care provided in our 2001-2002 multi-state validation sample. We selected a master list of validation variables that measure the care processes, management structure and general environment of a facility. The validation variables could be individual data elements collected in Administrative Survey (AS), Facility Walk Through (WAM) and Medicare Record Review (MRR) (referred to as items), or they could be any of 160+ validation scales derived from the items (referred to as scales) by the MegaQI Team for the validation analyses conducted (and initially reported) in 2002.

Selection of Validation Items and Scales

By assumption, the validation variables were divided into *preventive variables*, expected to capture policies or actions that facilities implement in advance, to minimize the emergence of problems, and *responsive variables*, expected to capture actions that facilities may use as they recognize that residents have ongoing or emerging problems. This taxonomy creates an expected pattern of the direction of association. On the one hand, preventive variables should be negatively associated with QMs that measure deterioration or no improvement, but they should be positively associated with QMs that measure improvement or no impairment. On the other hand, responsive variables should be positively associated with QMs that measure deterioration or no improvement, but negatively associated with QMs that measure improvement or no impairment.

We initially selected all the preventive items that had been used to validate either chronic or post-acute QMs in the MegaQI study. Overall, 42 preventive items were selected from the Administrative Survey and Facility Walk Through[3]. We selected the preventive items for PAC QM candidates using the same selection method employed in the MegaQI project. Responsive items were summary variables derived from the Medical Record Review (MRR), a structured medical record review. In the MegaQI project, for a given QM, only responsive items addressing the same area of care were used for the QM validation analysis; for all areas of care, the same nine responsive items were selected. Accordingly, we selected nine corresponding items for each PAC QM for our analysis. However, there were three exceptions. For three QMs, shortness of breath (RSP02), balance (BAL01), and range of motion (ROM01), no corresponding responsive items were collected through the MRR. Therefore, no responsive items were tested for the three PAC QMs.

Through an intensive process of meeting and discussion, the MegaQI Team specified over 160 validation scales, derived from the individual items in the Validation Database. These scales were designed to represent the composite constructs of care processes or practice patterns of nursing homes. From this master list, we selected validation scales that were significantly associated with any of the chronic or PAC QMs in the hypothesized direction in the MegaQI study[4]. The selected scales could be dichotomized, categorical, counts (e.g., 10+), or continuous. Overall, 45 preventive scales

[3] Source: Validation of Long-Term and Post-Acute Care Quality Indicators June 10, 2003, Appendix K. An excel file was created based on Appendix K to summarize the items used for QM validation. See H:\HSRE\AFTERSHOCK\PAC ANALYSES\items.xls.

[4] Source: Validation of Long-Term and Post-Acute Care Quality Indicators June 10, 2003, Appendix I. An excel file was created based on Appendix I to summarize all the validation scales tested in the validation of chronic or acute QMs. There are 160+ validation scales. To reduce the workload, only scales that are significantly associated with QMs in the hypothesized direction at least once were selected for PAC QM validation analysis. See H:\HSRE\AFTERSHOCK\PAC ANALYSES\val_scale_list.xls.

(designated with "+" signs) and five responsive scales (designated with "-" signs) were selected for use in the bivariate analysis.

After selecting scales and validation items, we calculated Pearson correlation coefficients between all the PAC QMs and selected validation items or scales. To be included as an explanatory variable in a validation model, an item or scale had to have a bivariate correlation with the QM that was statistically significant at $p < 0.10$, the sign of the correlation had to be consistent, both with the nature of the QM and with the classification of the item, and the item or scale had to be significantly correlated, in the right direction, with at least two QMs.

5.0 Findings

Using the selection criteria listed above, we identified 35 preventive and five responsive variables for the TCU-only sample, versus 54 preventive and eight responsive variables for all-facility sample. Upon this review of the bivariate results, the project team, with CMS, decided to use the refined "all-facility" sample of nursing facilities (as previously described in PAC Facility Selection) to report validation findings. This decision was based on the degree of similarity between the two samples in overall validation statistics. Additionally, the results from the refined "all facility" group can be generalized to a larger group of nursing facilities, both TCU and other, that admit and treat PAC residents.

Following procedures used in the earlier MegaQI validation process, three multivariate models were estimated for each PAC QM. One included only preventive variables, one included only responsive variables, and the third included both preventive and responsive variables. The multiple R's from the multivariable regression models were the main indications for us to judge the validity of a PAC QM.

In Table E-1 we summarize the results from the multivariate regression modeling. The rows of the table reference the individual QMs with sequence number, labels/names, their descriptions, and seven additional data elements. Columns four, five and six present the counts of significant, supportive validation variables for each QM, with separate counts for the number that fall under the preventive and responsive domains, and a final count of the total number of supportive validation elements for the indicator. Columns seven through nine provide the Multiple R correlation estimates of the relationship between the pool of significant and supportive validation variables and the quality measure. The last column in the table, labeled "Degree of Validity", provides the final assessment of the confidence one can have in the quality indicator at the end of this validation process. There are three possible classifications:

- Level I, Highest Validity, represents those quality indicators with the strongest support.
- Level II, Moderate Validity, achieved lesser support but are still considered to be valid.
- Level III, Not Validated, represent measures that failed to be supported in this analysis. In their current form, there is insufficient reason to believe that they provide a reasonable facility estimate for the quality problems they seek to address.

Below is a summary of our findings.

- Of the 26 PAC QMs tested, 19 had Level I validity, one had Level II and six had Level III.

- Of the three PAC QMs currently reported, DEL0X1 had Level I validity, PAI0X1 had Level I validity and PRU0X2 had Level III validity. Our results were similar to the

findings in the 2003 Validation Report of the MegaQI Project: in that report, the validity levels of these three QMs were Levels I, I and III, respectively

- Additionally, we compared our results to the validity levels of three QMs that were tested in all facilities in the MegaQI project, but were not public reported. The validity of CNT0X1, RSP0X1 and WAL0X1 were all Level I.[5]

- The validity of the QMs addressing the same care areas or conditions was often similar. For example, although ADL QMs have different denominators, numerators and exclusions, all achieved Level I validity.

- QMs with the same numerators and denominators often performed similarly in the validation analyses.

[5] For details see Validation Report, Appendix M. Our results were comparable to MegaQI validation results on all facilities without FAP. The QMs were named cnt0x, rsp0x and wal0x.

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
1	PAC-DEL0X1/ Currently reported	Percent of short-stay residents with delirium	5	3	8	0.53	0.47	0.66	I
2	PAC-DELOX2/ Covariate report	Percent of short-stay residents with delirium	5	3	8	0.53	0.46	0.66	I
3	PAC-PAIOX1 / Currently reported	Percent of short-stay residents who had moderate to severe pain	6	0	6	0.53	---	0.53	I
4	PAC-PAIOX2 / Covariate report	Percent of short-stay residents with pain	5	0	5	0.47	---	0.47	I
5	PAC-PRU0X1	Percent of short-stay residents whose pressure sores have not gotten better.	1	1	2	0.23	0.28	0.38	III
6	PAC-PRU0X2/ Currently reported	Percent of residents whose pressure sores have not gotten better.	0	0	0	---	---	---	III
7	PAC-PRU0X3 / Covariate report	Percent of short-stay residents whose pressure sores have not gotten better	1	1	2	0.25	0.24	0.36	III

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
8	PAC-ADL04 / HRCA, 2/2004	Percent of residents with improving level of ADL functioning	8	0	8	0.79	---	0.79	I
9	PAC-ADL05 / HRCA, 2/2004	Percent of residents who improve status on mid-loss ADL functioning (transfer, locomotion) or remain completely independent in mid-loss ADLs	11	0	11	0.80	---	0.80	I
10	PAC-ADL06 / HRCA,2/2004	Percent of residents who improve status on early-loss ADL functioning (dressing and personal hygiene) or remain completely independent in early-loss ADLs (ELADL).	7	0	7	0.68	---	0.68	I

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
11	PAC-MOD04/ HRCA,2/2004	Percent of residents who improve their mood or remain free from symptoms of depression (based on MDS Depression Rating Scale)	0	1	1	---	0.23	---	III
12	PAC-CAT03/ HRCA,2/2004	Percent of residents who do not have a catheter at 14-day assessment	0	0	0	---	---	---	III
13	PAC-PAI02/ HRCA,2/2004	Percent of residents who improve their pain status or remain free from pain, (based on the MDS Pain Scale)	9	1	10	0.55	0.25	0.58	I
14	PAC-RSP02/ HRCA,2/2004	Percent of residents who do not have shortness of breath	7	0	7	0.65	---	0.65	I

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
15	PAC-BAL01/ HRCA,2/2004	Percent of residents who improve their balance function or remain free from impairment in balance function between 5 and 14-day assessment	8	0	8	0.75	---	0.75	I
16	PAC-ROM01/ HRCA,2/2004	Percent of residents who improve their range of motion or remain free from impairment in ROM between 5 and 14-day assessment	0	0	0	---	---	---	III
17	PAC-ADL0X1/ Validation Report	Percent of residents who have not improved since admission	14	0	14	0.81	---	0.81	I
18	PAC-ADL0X2/ Covariate report	Percent of short-stay residents who have not improved since admission	11	0	11	0.70	---	0.70	I
19	PAC-CNT0X1/ Validation Report	Failure to Improve Bladder Incontinence	5	1	6	0.47	0.24	0.47	I

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
20	PAC-CNT0X2/ Covariate report	Failure to Improve Bladder Incontinence	5	0	5	0.49	---	0.49	I
21	PAC-RSP0X1/ Covariate Report	Percent of short-stay residents who have developed a respiratory infection or have not gotten better	5	0	5	0.57	---	0.57	I
22	PAC-RSP0X2/ Covariate Report	Percent of short-stay residents who have developed a respiratory infection or have not gotten better	4	1	5	0.48	0.24	0.52	I
23	PAC-WAL0X1 / Validation Report	Percent of residents who walk as well or better on day 14 as on day 5 of their stay.	8	0	8	0.72	---	0.72	I
24	PAC-WAL0X2 / Covariate report	Percent of short-stay residents who walk as well or better on day 14 as on day 5 of their stay	9	0	9	0.71	---	0.71	I

Table E-1: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
25	PAC-WAL02/ HRCA,2/2004	Percent of residents with improving level of locomotion functioning or who remained completely independent in locomotion functioning	5	0	5	0.66	---	0.66	I
26	PAC-CNT04 / Validation Report for Chronic	Prevalence of urinary tract infections	1	1	2	0.27	0.30	0.43	II

6.0 Discussion

As displayed in Table E-1, the majority (77 percent) of these "new" post-acute care quality measures appear to be valid, at either high (Level I) or mid (Level II) validity levels. A large set of measures across multiple domains of care provides multiple views of care quality, which will be beneficial to providers who seek to improve their quality of care processes. Given the size of the set, however, there will be a need to select among and prioritize the valid indicators for public reporting and other uses.

The care areas, and levels of functioning or health status represented within these domains are broad, and exceed the number of PAC QMs validated through the earlier MegaQI study. In the domain of clinical complexity, we include four separate clinical concepts for which at least one QM representing that concept has Level I or Level II validity: delirium; pain; bladder functioning; and respiratory functioning. We have also expanded the breadth of the functional status domain, by presenting not only improvement measures, which we believe can encourage positive practices among providers, but measures that address aspects of functioning not yet addressed and very relevant to the post-acute care population. For example, two new measures presented here capture improvements in "early-loss" and "mid-loss" (transfer and locomotion) ADLs.

Change in status QMs, especially those that capture improvement in function and clinical condition, are important for monitoring post-acute care, as the primary focus of post-acute care is on the restoration of residents' function. Two of the three publicly reported PAC QMs are cross-sectional QMs (prevalence), though the original MegaQI validation study did find four valid PAC incidence (change in status) measures. In our current analyses, we validated both cross-sectional and incidence measures. Nineteen of the 26 PAC measures are incidence measures; 14 incidence measures total were found to have Level I or Level II validity.

As has long been the stance of this project team, we favor risk-adjusted measures over unadjusted, in most cases, and hope that CMS will consider some of the validated measures adjusted with RUG-based covariates, since they should better capture casemix-related risk for those outcomes they are associated with. In this set of 26 measures, there are five concepts or care areas represented by multiple models or forms of the QM. For example, there are three pain measures, all found to have high (Level I) validity. Two of the pain measures are prevalence measures, and one is a change in status measure (residents improve pain status or remain free from pain). Of the two prevalence measures, one (PAI0X1) is unadjusted, while the other (PAI0X2) is risk-adjusted with resident-level covariates (CPS, Hip Fracture and Alzheimer's/Other Dementia).

Table E-2 categorizes each QM by the factors discussed above (e.g., quality domain, prevalence vs. incidence, risk-adjustment).

Table E-2: QM Characteristics

Domain	Quality Measure	Concept Represented	Type of Measure	Risk Adjustment
Clinical Complexity	PAC-DEL0X1	Delirium	Prevalence	Prior residential history
	PAC-DEL0X2	Delirium	Prevalence	None
	PAC-PAI0X1	Pain	Prevalence	None
	PAC-PAI0X2	Pain	Prevalence	CPS, Hip Fracture, Alzheimer's/Other Dementia)
	PAC-PAI02	Pain	Incidence, Improvement	CPS, Hip Fracture
	PAC-PRU0X1	Pressure Sores	Incidence, Decline	None
	PAC-PRU0X2	Pressure Sores	Incidence, Decline	Unresolved pressure sore, bed mobility, bowel incontinence, DM or PVD, low BMI
	PAC-PRU0X 3	Pressure Sores	Incidence, Decline	R_CLN, R_CMI, R_ADL
	PAC-CAT03	Bladder Function	Prevalence	R_ADL, R_CLN
	PAC-CNT0X1	Bladder Function	Incidence, Decline	None
	PAC-CNT0X2	Bladder Function	Incidence, Decline	R_CMI, R_ADL
	PAC-CNT04	Bladder Function	Prevalence	None
	PAC-RSP0X1	Respiratory Function	Incidence, Decline	Asthma, Emphysema/COPD
	PAC-RSP0X2	Respiratory Function	Incidence, Decline	R_CLN
	PAC-RSP02	Respiratory Function	Prevalence	Emphysema/COPD
Functional Status	PAC-ADL04	ADLs	Incidence, Improvement	CPS
	PAC-ADL0X1	ADLs	Incidence, Decline	No prior residential history
	PAC-ADL0X2	ADLs	Incidence, Improvement	CPS
	PAC-ADL05	ADLs	Incidence, Improvement	CPS, R_ADL
	PAC-ADL06	ADLs	Incidence, Improvement	CPS, R_ADL
	PAC-BAL01	Balance	Incidence, Improvement	CPS, R_ADL
	PAC-ROM01	Range of Motion	Incidence, Improvement	R_ADL, hip fracture
Mobility	PAC-WAL0X1	Walking	Incidence, Improvement	None

Table E-2: QM Characteristics

Domain	Quality Measure	Concept Represented	Type of Measure	Risk Adjustment
	PAC-WAL0X2	Walking	Incidence, Improvement	R_ADL, R_CMI, CPS
	PAC-WAL02	Locomotion	Incidence, Improvement	CPS, R_ADL
Mood	PAC-MOD04	Depression	Incidence, Improvement	None

7.0 Conclusions and Next Steps

Together with colleagues from HRCA, Abt Associates accomplished the two principal objectives of this project: 1) to acquire, test and deliver to CMS a documented copy of the Validation Database, and 2) to identify and validate additional candidate PAC QMs. We also made substantial progress in expanding the number of valid PAC QMs.

- We were able to demonstrate high or moderate levels of validity for 20 additional PAC QM measures.
- Our findings did not substantially change the validity determinations for the three PAC QMs that are now publicly reported.
- Compared to the 2002 analyses, which validated PAC QMs on TCUs only, we used a more broadly representative sample of nursing facilities.

Though we hesitate to recommend any one measure over another, we do encourage CMS to consider the potential benefit of having multiple measures of a given clinical or functional concept, in order to capture the multidimensionality of nursing facility care quality. In this vein, it may be beneficial to publicly report an incidence and prevalence pain measure, for example, or an improvement and decline measure for the sum of ADLs presented in the ADL Long Form summary scale (e.g., ADL04 and ADL0X2). As stated previously, we do recommend those QMs that are risk-adjusted, over others that are not.

At the end of the MegaQI project, we identified additional analyses that CMS might consider for extending or refining the current QM system. Here, we note additional analyses that might be undertaken to expand upon and contribute to this body of work:

- QMs might be developed and tested for "special populations" admitted for post-acute care, such as dementia residents or residents with special mental health needs.
- Sensitivity tests for validity might be conducted. The MegaQI Team assigned levels of validity based on statistical measures of association. In some cases, these measures of association varied narrowly around thresholds for accepting or rejecting a QM. Further exploration of the validity of the current set of measures, using differing methods, may yield useful information regarding the relationship of the measures to facility process measures of quality.

- The use of facility-level risk adjustors continues to show promise for producing QMs with high discriminatory power. Further work performed by HRCA on a direct adjustment method for capturing facility casemix also shows promise. These issues warrant further exploration and eventual adoption for public reporting.

- One large task of the MegaQI contract, that of establishing benchmarks for ranking providers on the quality of care they provide, was curtailed due to the more pressing needs of public reporting.

- The current Administration has shown interest in "pay for performance" systems, and on their implementation, possibly on a demonstration basis, in nursing homes, home health agencies and physicians offices (hospital pay for performance demonstrations and voluntary programs are now underway). In advance of a mandated demonstration, CMS might undertake an assessment of what changes, if any, in the current system of quality measures might be appropriate to support nursing home pay for performance (for example, could the current QMs provide the basis for a scoring system similar to the one currently in place for the Premier hospital pay for performance demonstration?).

1.0 Introduction

1.1 Scope of Work

In May 2004, the RAND Corporation subcontracted with Abt Associates Inc. to address several tasks related to work done on the nursing home project "Development and Validation of Indicators and Measures of Quality and Appropriateness of Services Rendered in Post-acute and Long-term Care Settings" (the MegaQI Project).[6] The scope of work for this subcontract included the following:

1. Refresh CMS' publicly reported nursing home quality measure (QM) data by providing national facility and resident-level QM files in June and September 2004.[7]

2. Document and deliver to CMS the "Validation Database", developed by the MegaQI Team. This database included raw data and measures constructed from primary data collected in 209 nursing homes, together with statistical programs used to validate the MegaQI Team's recommended QMs as reported in "Validation of Long-Term and Post-Acute Care Quality Indicators (September 2002) (Validation Report). This task also included re-creation of Appendix M from the Validation Report, a table of validation results for PAC QMs that had been specified in two versions: unadjusted and adjusted by the Facility Admissions Profile (FAP).

3. Review, analyze and recommend post-acute care (PAC) quality of care measures from existing chronic and PAC measures. This task included:

    - Exploring the feasibility of expanding the number of valid PAC QMs, by reviewing and testing the PAC measures originally considered by the MegaQI Team, a set of PAC QMs proposed by HRCA in a memo to CMS in February 2004, and certain chronic QMs that "have face validity for post acute residents."

    - For some of the PAC QMs, applying risk adjustment, as appropriate.

    - Testing the validity of the expanded list of PAC QMs on two components of the 209 nursing homes: a) transitional care units (TCUs), hospital affiliated nursing facilities that were used to validate all PAC QMs in the 2003 Validation Report, and b) a larger sample that would include TCUs and other nursing facilities.

[6] Abt Associates' subcontract was written under the RAND Corporation's prime contract with CMS, "Development and Validation of a Revised Nursing Home Assessment Tool: MDS 3.0" The subcontract was called "Quality Measures (QMs) Addendum' (a revision to Task 4 of the prime contract: Quality Measurement Development and Refinement). The MegaQI contract was called "Development and Validation of Indicators and Measures of Quality and Appropriateness of Services Rendered in Post-acute and Long-term Care Settings" CMS Contract No: 500-95-0062/T.)/ No. 4.

[7] Abt Associates programming staff completed the first QM update, providing CMS access to both resident and facility-level files for the 10 QMs first publicly reported by CMS (and in continued use for quality improvement initiatives by nursing facilities and Quality Improvement Organizations) and the 14 QMs that were publicly reported starting in January 2004, based on recommendations from the National Quality Forum (NQF). Abt did not complete the second round of updates. CMS requested the Iowa Foundation for Medical Care (IFMC) to take over this task, beginning with the September update.

1.2 Organization of the Report

Section 2.0 of this report briefly reviews CMS objectives for this project, and summarizes the major accomplishments and findings. Section 3.0 describes the Validation Database and the collaborative work done by Abt Associates and HRCA staff to document and prepare it for delivery to CMS. Section 4.0 describes the PAC QM refinement and expansion task in terms of four steps:

- Selecting a set of PAC QMs for validation
- Specifying and writing code for processing the recommended QM set
- Selecting nursing facility samples for the validation analyses
- Conducting the validation analyses

Section 5.0 reports findings from the PAC QM validation analyses. Section 6.0 discusses and interprets these findings, and presents recommendations. Section 7.0 concludes the report and suggests possible next steps.

2.0 Purpose and Overview

2.1 Purpose of the Project

CMS' principal objective for this project was to learn whether or not it would be feasible to expand the set of valid PAC QMs, using data and methods applied in the MegaQI 2003 Validation Report. In this report, the MegaQI Team recommended six PAC QMs. Four were in the category Level I ("highest validity") and two were in Level II ("moderate validity"). In contrast, 29 chronic care QMs were recommended, having achieved either Level I or II.

Of the 10 QMs initially reported by CMS, four were PAC QMs. Some were unadjusted, and some adjusted by the Facility Admission Profile (FAP):

- Percent of short-stay residents with pain
- Percent of short-stay residents with delirium without FAP
- Percent of short-stay residents with delirium with FAP
- Percent of short-stay residents who walk as well or better on day 14 as on day 5 of their stay, with FAP

Subsequently, the National Quality Forum (NQF) conducted a review of the MegaQI team's proposed QMs. After completing this review, the NQF recommended eleven chronic and post-acute care (PAC) measures for the public reporting pilot project (restraints, weight loss, new infections, pain management prevalence, pressure sore prevalence, antipsychotic use, ADL decline, rehospitalization (PAC), inadequate pain management (PAC), delirium (PAC) and walking improvement (PAC)). Ten measures were subsequently reported, since the rehospitalization measure developed by the MegaQI Team was not adequately developed, constructed or tested.

The NQF also considered the proposed risk adjustment methodology, including the somewhat controversial FAP adjustor. Additional work was performed by the MegaQI Team on resident level risk adjustment, since this appeared to be of strong interest to NQF Steering Committee members and to the industry. However, the final set of measures endorsed on October 3, 2003 by the NQF for

national public reporting did not include the risk-adjusted QMs proposed by the MegaQI Team in December 2002. The FAP-adjusted QMs were also dropped. Current publicly reported PAC QMs are:

- Percent of short-stay residents with delirium;
- Percent of short-stay residents who had moderate to severe pain; and
- Percent of short-stay residents whose pressure sores have not gotten better.

CMS and others have expressed concern about the dearth of valid PAC QMs. Medicare pays for much of the post-acute care provided by nursing facilities. Therefore, measuring facilities' performance in caring for short-stay residents is clearly an important issue for CMS and for persons who make decisions about post-acute placement (discharge planners, social workers, residents and family members). Adding to the number of valid QMs would expand the scope of the quality measurement system, to embrace more of the post-acute care that facilities provide.

To support this additional validation work, CMS wanted to make sure that:

- The Validation Database created for the MegaQI contract would be accessible by programmers at CMS.
- The Validation Database was sufficiently documented to support additional analytic work.
- CMS, and Abt and the Hebrew Rehabilitation Center for the Aged (HRCA) were in agreement, both on the processes employed to combine QMs with measures based on data collected from nursing facilities in validation analyses and on the outcomes of those analyses.

Thus, CMS asked Abt Associates to fully understand and document the architecture of the Validation Database, the code used to create validation scales and elements, the specifications and code for all proposed PAC QMs, procedures for selecting facilities for validation analyses, and statistical routines used to validate the proposed QMs. At CMS' request, Abt worked with senior researchers from HRCA, Drs. John Morris and Richard Jones, who led the design and implementation of the data collection and analyses that led to recommendations in the 2003 Validation Report. As a tangible way to demonstrate full understanding of the validation process, CMS asked Abt to replicate the statistics in Appendix M of the Validation Report. This table reported multiple R statistics from validation models, run both on the transitional care units (TCUs) and full sample of nursing facilities from which the MegaQI Team collected data, and assigned validity levels based on these statistics.

3.0 Description of Validation Database and Work Completed to Submit as Deliverable

CMS asked Abt Associates to test and submit to CMS a copy of the MegaQI Validation Database. In order to understand fully the database and programs used for validation, CMS also asked Abt to replicate Appendix Table M of the 2003 report "Validation of Long-Term and Post-Acute Care Quality Indicators" (the Validation Report). Appendix Table M presented measures of multivariate correlation (R-statistics) and validity levels for seven Post-Acute Care Quality Measures (PAC QMs).

The first section of this chapter describes the Validation Database. The second section briefly describes the process of acquiring and testing the data files and programs of the database, including replication of Appendix Table M.

3.1 Description of the Validation Database[8]

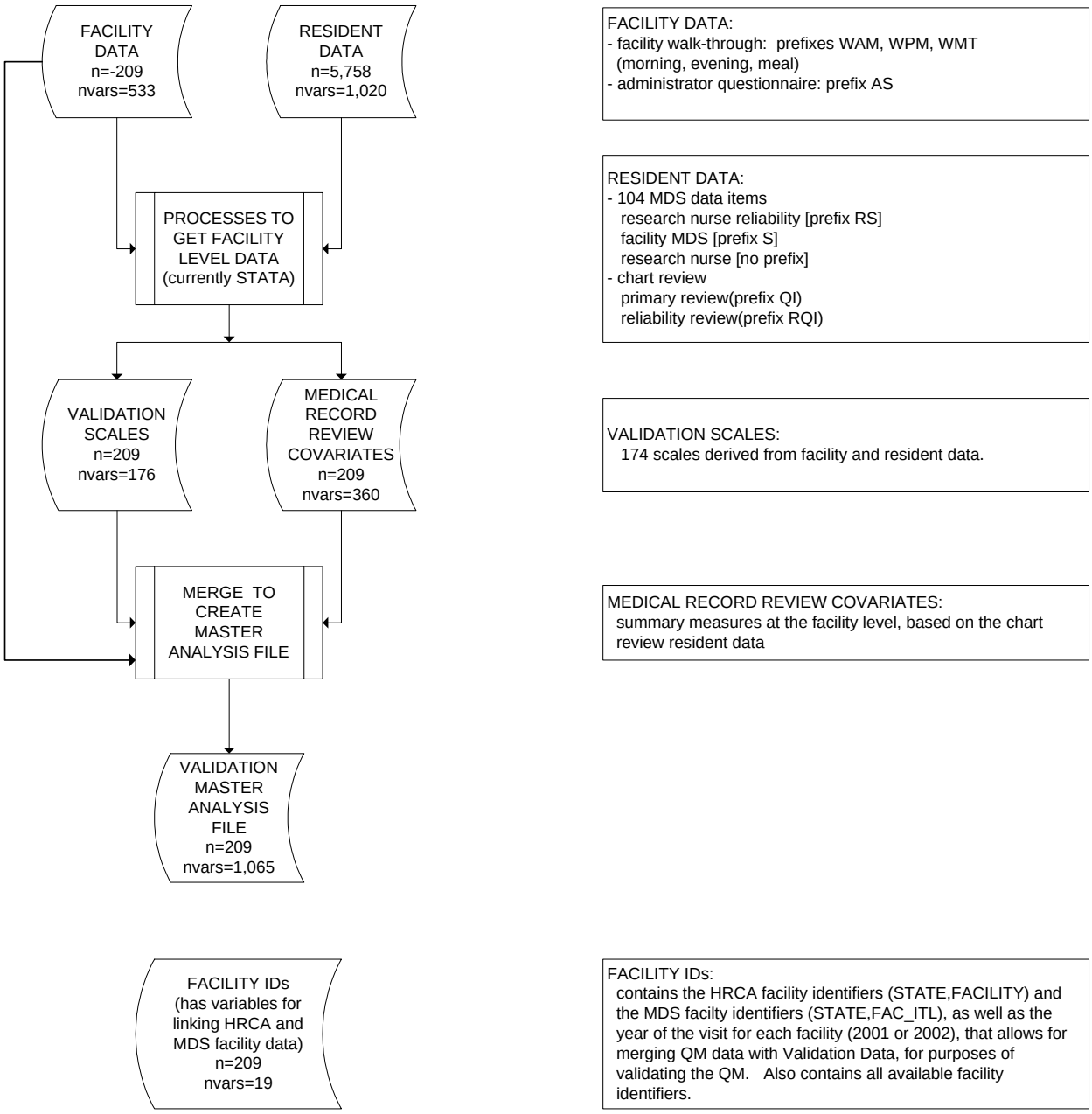
The objective of the MegaQI validation study was to demonstrate a correlation between measures of nursing facility processes of care and resident-based outcome measures of presumed quality of care (QMs). Facility-level structure and process of care measures were collected directly by trained research staff using structured interview protocols, medical record review, and a questionnaire administered to key facility informants (administrators, nursing directors). The final analytic sample consisted of 209 freestanding and hospital-based facilities located in six states: California, Illinois, Missouri, Ohio, Pennsylvania and Tennessee. Medical record review data were based on a sample of residents within each facility, with a goal of reviewing 30 residents' charts per facility. The total resident sample included in the on-site field review comprised some 5,758 residents.

Data collection occurred at both the facility and resident levels. Facility-level primary data were linked to QMs developed from MDS data with a crosswalk file of facility ID variables. For the resident-level file, unique records (and residents) are identified by state and facility ID.

The structure of the Validation Database is depicted in Figure 1.

[8] Based on a memo to Abt Associates Inc. written by Richard N. Jones, Margaret Bryan, William McMullen, Shirley Morris and John N. Morris of the Hebrew Rehabilitation Center for Aged Research and Training Institute "MegaQI Validation Study: Data Documentation" (April 16, 2004). This memo was repeated in its entirety in Abt Associates Inc. "MegaQI Validation Database Documentation" (August 18, 2004), submitted to CMS with a CD copy of the database.

Figure 1: QM Validation Database



3.2 Data Files

RESIDENT.SAS7BDAT

This file contains all of the data collected at the resident level, including both MDS reliability data and data from a structured medical record review. The 104 MDS items collected on residents

constitute a subset of the items on the full MDS; most of these items were required in the construction of one or more of the QMs under evaluation.

The structured medical record review directed nurse assessors to review each sampled resident's chart for evidence of care processes in 21 areas: a) cognition, b) communication, c) delirium, d) mood, e) behavior, f) ADL improvement, g) ADL decline, h) mobility, i) falls, j) antipsychotic drugs, k) pain, l) restraints, m) feeding tubes, n) nutrition, o) catheter, p) bladder incontinence, q) bowel incontinence, r) infections, s) pressure sores, t) burns, and u) little or no involvement in activities. Four main areas or domains were addressed for each clinical condition, covering 1) assessment, 2) documentation, 3) response to change in resident's condition and 4) care planning. Each domain of investigation then had between three and five sub-items.

FACILITY.SAS7BDAT

This file contains all of the data collected at the facility-level. There are two categories of data in this file: **Facility Walk-Through** items and **Administrative Survey** items.

Facility Walk-Through. The aim of the Environmental Walk Through/Resident Observation was to gain an overall understanding regarding whether the facility was "resident-centered", what the "feel" of the facility was, and what the nature of staff interactions with residents were. A series of general environmental measures were employed to describe the responsiveness of the milieu to resident strengths, needs, and problems that included general care environment measures (e.g., nature of physical environment, communication strategies, environmental manipulation and resident interactions with staff). These measures were collected through assessment, surveillance, and observation of staff technique. The data collectors on site recorded their observations three times per day at approximately 10:00 a.m., lunchtime, and 2:00 p.m. to obtain a comprehensive picture of the facility care environment.

Administrative Survey. The Administrative questionnaire included questions regarding staffing, residents, specialists, clinical communication protocols, care planning and training and orientation of staff.

FACILITY_IDS.SAS7BDAT

This facility-level SAS database contains the identifiers necessary for matching a record in the validation database (using state and facility identifiers) with a record from the quarterly QM data.

V_SCALES.SAS7BDAT

Based on discussions of clinical expert panels, the MegaQI Validation Study Steering committee developed, a priori, a set of measures believed to predict nursing home QM scores. Each scale is defined in Appendix F of the 2003 Validation Report. These variables, computed from MDS, Chart Review, Walk-through, and Administrative Survey data, were derived from data in the Resident and Facility files described above.

MRR_COVARIATES_ ALL.SAS7BDAT

This file contains summary data from the medical record review (MRR), for all facilities. Data are aggregated from the resident level, where the variables are categorical variables (= 0 or 1), to the facility level, where the calculated mean is a proportion.

ANALYSIS.SAS7BDAT

This facility-level file includes all of the variables that are available in the Validation Database for use in attempting to validate a QM. These include state and facility identifiers, and variables constructed from the primary data collection components.

3.3 Acquiring, Reviewing and Preparing the Validation Database

Beginning in April 2004 and continuing into August, Abt Associates and HRCA analysts and programmers worked closely together to document the Validation Database for delivery to CMS. As requested, in June, Abt programmers first tried to replicate findings reported in Appendix Table M of the Validation Report. In addition, attempts were made to replicate Table 2 from the same report. [9] However, initial efforts to replicate these tables were unsuccessful. Later, after sustained effort and discussion, Abt programmers were able to replicate the statistics in these two tables. In August 2004, once the tables were successfully replicated, the Abt team submitted the data files and programs in the Validation Database, together with full documentation, to CMS.

Having initially failed to replicate Appendix Table M and Table 2, the Abt team explored with HRCA possible reasons for this failure. During June and July, Abt and HRCA programmers maintained an active dialog, researched model specifications and variable definitions, shared code and ran parallel analyses. After a lengthy process of discovery, the following sources of the initial failure were uncovered:

- The data Abt received from HRCA in June 2004 were slightly different from the data used in the 2002 validation analyses. After Abt submitted the 2003 Validation Report to CMS, HRCA received additional data from some of the sampled nursing facilities. In fact, one more facility than the 2003 sample provided complete data after the report submission. These additional data were subsequently added to the Validation Database.

- The Abt/HRCA Team had to reconstruct the methodology used to estimate the earlier multivariate validation models. Initially, Abt estimated regressions one by one, from raw data. However, HRCA had obtained the multiple R's needed for assessing validity levels from a regression model based on a saved covariance or correlation matrix as opposed to raw data. Tests demonstrated that each method yields slightly different results, because each accounts for missing values in a different way.

- Different validation models were used to generate statistics for different samples of nursing facilities in Appendix M. Appendix M presented results for 12 regressions:

  - Two samples: transitional care units (TCUs) and all nursing facilities.
  - Two adjustment models: FAP and no FAP.
  - Three basic model specifications: preventive elements only, responsive elements only, both preventive and responsive elements.

- For each QM, the validation model for the TCU sample could be (and generally was) different from the model for all the nursing facilities.

[9] Table 2 presented validation findings, including estimated levels of validity and the statistics on which these levels were based.

After reviewing all these issues in late June, HRCA produced and delivered to Abt a copy of the Validation Database that had been used in analyses for the 2003 Validation Report. With this database, and a more complete understanding of the estimation methodology, Abt programmers were able to replicate the results in Appendix Table M and Table 2. On August 18, 2004, Abt submitted to CMS a CD containing the full Validation Database, together with documentation.

4.0 Methods

4.1 Selection of PAC QMs for Validation

The selection of PAC quality measures for the current project began from an initial list of 29 candidate PAC QMs. These included publicly reported measures, measures developed and tested for validation by the MegaQI project but not publicly reported, new measures proposed by HRCA in a memo to CMS (January 2004), and chronic care QMs that the team considered relevant to post-acute care. A further review of these measures is provided below.

We reviewed all of the PAC measures developed through the MegaQI project. Some of these had already been tested for validation, as reported in the Validation Report.[10] Others had been selected for subsequent exploratory covariate analyses summarized in a December 2002 MegaQI report.[11] From these sets of measures, we selected 15 PAC and three chronic care (CC) QMs. Of the PAC QMs, three are publicly reported, including prevalence of short-stay residents with delirium (PAC-DEL0X1), pain (PAC-PAI0X1) and pressure ulcer (PAC-PRU0X2). Five of the selected PAC QMs were tested for validation in the MegaQI project but were not chosen for public reporting. Seven were included in the covariate analyses. The remaining three QMs were originally designed to measure the quality of chronic care; these include the prevalence of cognition worsening, urinary tract infection and daily use of physical restraints. The research team believed that these QMs might also be relevant to the care delivered to post-acute residents of nursing facilities. The validity of the three chronic care QMs had been tested (for chronic care residents) in the earlier validation effort. The purpose of the current validation effort was to test their validity for short-stay (or PAC) residents only.

In 2004, HRCA developed a new set of PAC QMs, not previously validated or analyzed by the MegaQI project. From HRCA's candidate PAC QMs, we selected eleven for further testing. These QMs focus primarily on functional abilities, and on functional improvements between post-acute admission and the 14-day MDS assessment. This project represents the first attempt to validate the newly developed QMs with a large data set collected from multiple facilities.

Initially, we had also considered attempting to validate a measure of re-hospitalization. This measure had been discussed and discarded by the MegaQI team for a variety of conceptual and practical reasons. Most important, creating and testing a useful measure of re-hospitalization requires merging Medicare claims with MDS assessment data, a task beyond the scope of the current project. Our

[10] Morris, J. et al. Validation of Long-Term and Post-Acute Care Quality Indicators: Draft Final Report. June 10, 2003.

[11] Kidder, D, et al. MegaQI Covariate Analysis and Recommendations: Identification and Evaluation of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings December 20, 2002

understanding is that there are efforts underway, led by the Agency for Healthcare Research and Quality (AHRQ), to develop a re-hospitalization measure.[12]

Once the initial set of 29 measures was decided upon, the technical specifications were drafted, SAS code written, and the measures were processed against the MDS repository data for the period 2001-2002. We then examined the distribution of the PAC QMs and found three to fail our criteria for selection (insufficient or excessive prevalence). The three QMs dropped from further analysis were:

- COM02—No decline in communicative function
- RES01—Restraints (physical) used daily, prevalence
- COG01—Cognition worsening

The average incidence of COM02 was 0.97 (interquartile range: 0.95-1.00). The average prevalence of RES01 was 0.05 (interquartile range: 0.00-0.07). The average incidence of COG01 was 0.05 (0.00-0.08).[13] Although these QMs address important areas of care, and perform well when applied to chronic care residents, the between-facility variation in these QMs is not great enough to allow us to separate good and bad care processes. Thus, these three measures, as defined here, are not meaningful in the measurement of post-acute care quality. The remaining 26 PAC QMs thus comprised the final list to be moved forward to validation testing (see Appendix 1 for the final QM specifications).

4.2 Processing the Recommended QMs

QM data file construction proceeded in two steps. First, we computed the recommended QMs on all MDS records for the US for four target quarters. Each QM could have at most two versions: adjusted (using regression techniques, if resident-level covariates were recommended) and unadjusted (except for exclusions from the numerator or denominator). Then we merged a file with the unadjusted and adjusted QMs to data from the Validation Database for all facilities in the Validation Sample (a subset of the 209 facilities providing data for the 2003 Validation Report -- see section 4.3 below).

Construction of the Quality Measures

We obtained all MDS records from CMS for the period 2000 through 2002 and used them to construct the PAC QMs for four target quarters: 2001Q3, 2001Q4, 2002Q1 and 2002Q2.[14]

We employed two approaches for risk adjusting the QMs. First, we excluded residents whose outcomes were not under facility control or the outcome may have been unavoidable. For PAC QMs that measure delirium, ADL, balance, mobility, incontinence and walking, we excluded three groups of residents from the denominators and numerators:

- Residents in a coma (B1 = 1) or comatose status unknown (B1 = missing) on 14-day assessment;

[12] Telephone conversation with Robert Godbout, Stepwise Systems, September 2004.

[13] Source: H:\HSRE\AFTERSHOCK\PAC ANALYSES\qmstats_qt_states.xls (created on 11/10/2004, shared with HRCA and CMS on 11/10/2004). The statistics are from the worksheet qmstat1a, row "all".

[14] We did not use the processed data sets from MegaQI project because the covariates of the newly developed QMs were not included in the already-processed data set.

- Residents with end stage disease (J5c=checked) or end stage disease status unknown (J5c = missing) on 14-day assessment; and

- Residents enrolled in hospice (P1ao = checked) or hospice status unknown (P1ao = missing) on 14-day assessment.

Second, for some QMs, we adjusted for risk using resident-level covariates that have been found to increase the risk for an outcome. For instances in which covariates had already been specified for previously-developed QMs, "original" covariates were maintained (e.g., covariates intended to measure residents' physical, social and cognitive function and clinical condition and diagnosis indicators, from items in Sections I and J of the MDS). We also used indices and scales derived from the Resource Utilization Group-III (RUG-III) system, now used to adjust Medicare payments to nursing facilities -- these included the Nursing Case Mix Index (CMI) used for Medicare; and scales created from the RUG CMI model (scales for Clinically Complex and Late Loss ADL); and the Cognitive Performance Scale (CPS). Specifications for these covariates may be found in Appendix 2.

An "unadjusted QM" is calculated as the total number of residents for whom the QM is triggered, divided by the total number of residents who are at risk. An "adjusted QM" is calculated in three steps:

- First, we calculate the "observed QMs" which is identical to an unadjusted QM, with the added exclusions of residents with missing values on any covariates.

- Second, we calculate the "expected QMs" at the facility level. The denominator of an expected QM is the same as that for an observed QM. However, the numerator is the predicted probability that a resident had an event triggering the QM, given that resident's covariate values and the estimated coefficients from a prediction model.

$$\text{Predicted Probability}_{ij} = \frac{1}{1+\exp[-(\sum \beta * \text{covariates}_{ij})]},$$

  where *i* indexes resident and *j* facility, β's are the regression coefficients we calculated using the MDS data from a year prior to the target quarter from a 10 percent randomly sampled set of facilities.

- Third, we calculate the adjusted QM for each facility using the equation below:

$$\text{Adjusted QM}_j = \frac{1}{1+\exp\{-[\text{logit}(\text{QM}_{\text{obs},j}) - \text{logit}(\text{QM}_{\text{exp},j}) + \text{logit}(\text{QM}_{\text{obs}})]\}},$$

where $\text{logit}(\text{QM}_{\text{obs},j})$ and $\text{logit}(\text{QM}_{\text{exp},j})$ are the observed and expected QMs for the *j*th facility, respectively; $\text{logit}(\text{QM}_{\text{obs}})$ is the observed QM score across all facilities in US.[15]

[15] For details, see Abt Associates Inc., National Nursing Home Quality Measures: Users Manual. January 2004, (V1.1)

Distribution of the Calculated QMs

After constructing and processing the PAC QMs, we examined their distributions. The distributions of the three currently publicly reported PAC QMs (DEL0X1, PAI0X1 and PRU0X2) were similar to the values presently publicly reported (Table 1).

Newly developed QMs also had reasonable values. Families of QMs representing the same clinical or functional areas (e.g., pressure ulcers) had similar distributions regardless of their respective risk-adjustment models. We further compared the values of the QMs across the four quarters. Over 9,000 facilities had QMs calculated for each of the four target quarters. We found very small fluctuations in average QMs over time; and the magnitude of these fluctuations was compatible with trends that we have observed and reported on for the publicly reported QMs. [16]

To provide some context for our validation efforts, we also compared the QMs for all nursing facilities in the six sampled validation states (from which the sample of 209 were selected). There was some variation in the QMs across states, which is also compatible with previous findings.[17] These findings demonstrate face validity and the potential utility of the set of QMs proposed for further analysis.

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-DEL0X1 / Currently reported	CA	663	0.029	0.058	0.000	0.000	0.039
	IL	461	0.041	0.062	0.000	0.023	0.057
	MO	218	0.061	0.078	0.000	0.037	0.095
	OH	598	0.048	0.053	0.000	0.035	0.074
	PA	570	0.042	0.051	0.000	0.029	0.063
	TN	240	0.046	0.067	0.000	0.025	0.057
PAC-DEL0X2 / Covariate report	CA	663	0.030	0.060	0.000	0.000	0.039
	IL	461	0.043	0.065	0.000	0.024	0.057
	MO	218	0.064	0.080	0.000	0.040	0.095
	OH	598	0.050	0.056	0.000	0.037	0.074
	PA	570	0.045	0.054	0.000	0.031	0.063
	TN	240	0.049	0.070	0.000	0.026	0.057
PAC-PAI0X1 / Currently reported	CA	663	0.286	0.157	0.162	0.280	0.386
	IL	461	0.236	0.154	0.118	0.207	0.320
	MO	218	0.296	0.147	0.191	0.290	0.389
	OH	598	0.283	0.140	0.185	0.268	0.364
	PA	570	0.233	0.124	0.143	0.214	0.313
	TN	240	0.281	0.147	0.174	0.275	0.384

[16] For example, see D. Kidder, L. Hadden and B. Bell. Analysis of Q4 2003 to Q1 2004 National NQF-Recommended QM Data: Revision to the June 23, 2004 memo. Memo to Zhoowan Jackson. July 22, 2004

[17] Idem.

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-PAI0X2 / Covariate report	CA	663	0.301	0.147	0.203	0.302	0.386
	IL	461	0.235	0.146	0.126	0.221	0.320
	MO	218	0.294	0.141	0.202	0.298	0.389
	OH	598	0.296	0.140	0.196	0.288	0.364
	PA	570	0.240	0.119	0.151	0.230	0.313
	TN	240	0.304	0.145	0.198	0.300	0.384
PAC-PRU0X1	CA	663	0.294	0.146	0.189	0.275	0.383
	IL	461	0.214	0.120	0.130	0.195	0.280
	MO	218	0.187	0.103	0.107	0.179	0.250
	OH	598	0.218	0.105	0.140	0.208	0.273
	PA	570	0.235	0.101	0.163	0.227	0.290
	TN	240	0.233	0.119	0.157	0.225	0.289
PAC-PRU0X2 / Currently reported	CA	663	0.263	0.129	0.169	0.242	0.383
	IL	461	0.221	0.114	0.140	0.212	0.280
	MO	217	0.194	0.105	0.118	0.186	0.250
	OH	598	0.217	0.104	0.144	0.200	0.273
	PA	570	0.224	0.097	0.155	0.213	0.290
	TN	240	0.214	0.108	0.143	0.211	0.289
PAC-PRU0X3 / Covariate report	CA	663	0.282	0.134	0.190	0.267	0.383
	IL	461	0.227	0.118	0.151	0.217	0.280
	MO	218	0.198	0.100	0.121	0.194	0.250
	OH	598	0.202	0.099	0.135	0.188	0.273
	PA	570	0.217	0.094	0.150	0.204	0.290
	TN	240	0.221	0.111	0.145	0.212	0.289
PAC-ADL04 / HRCA, 2/2004	CA	663	0.542	0.189	0.414	0.555	0.680
	IL	461	0.448	0.225	0.261	0.444	0.645
	MO	218	0.429	0.206	0.289	0.434	0.580
	OH	598	0.424	0.179	0.298	0.415	0.565
	PA	570	0.405	0.151	0.304	0.401	0.514
	TN	240	0.441	0.190	0.310	0.450	0.568
PAC-ADL05 / HRCA, 2/2004	CA	663	0.444	0.157	0.346	0.444	0.550
	IL	461	0.374	0.179	0.238	0.366	0.527
	MO	218	0.378	0.166	0.265	0.377	0.514
	OH	598	0.361	0.149	0.259	0.358	0.464
	PA	570	0.346	0.136	0.257	0.337	0.433
	TN	240	0.369	0.157	0.267	0.366	0.473

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-ADL06 / HRCA,2/2004	CA	663	0.328	0.160	0.211	0.317	0.438
	IL	461	0.274	0.179	0.128	0.249	0.436
	MO	218	0.264	0.157	0.149	0.253	0.379
	OH	598	0.246	0.145	0.138	0.225	0.355
	PA	570	0.239	0.125	0.145	0.224	0.313
	TN	240	0.242	0.154	0.127	0.219	0.322
PAC-COM02 / HRCA,2/2004 (dropped)	CA	663	0.972	0.039	0.957	0.983	1.000
	IL	461	0.974	0.036	0.957	0.987	1.000
	MO	218	0.968	0.044	0.950	0.989	1.000
	OH	598	0.977	0.032	0.962	1.000	1.000
	PA	570	0.944	0.061	0.917	0.960	1.000
	TN	240	0.970	0.038	0.955	0.982	1.000
PAC-MOD04 / HRCA,2/2004	CA	663	0.726	0.194	0.633	0.766	0.867
	IL	461	0.668	0.236	0.536	0.722	0.843
	MO	218	0.725	0.195	0.600	0.768	0.875
	OH	598	0.501	0.237	0.306	0.513	0.684
	PA	570	0.708	0.196	0.595	0.744	0.862
	TN	240	0.700	0.199	0.567	0.723	0.864
PAC-CAT03 / HRCA,2/2004	CA	663	0.231	0.117	0.141	0.223	0.311
	IL	461	0.204	0.103	0.128	0.192	0.256
	MO	218	0.219	0.124	0.126	0.206	0.250
	OH	598	0.186	0.100	0.112	0.172	0.265
	PA	570	0.169	0.093	0.100	0.156	0.250
	TN	240	0.197	0.099	0.122	0.178	0.274
PAC-PAI02 / HRCA,2/2004	CA	663	0.498	0.160	0.385	0.482	0.652
	IL	461	0.512	0.157	0.409	0.504	0.624
	MO	218	0.436	0.161	0.327	0.439	0.542
	OH	598	0.437	0.128	0.351	0.435	0.537
	PA	570	0.484	0.116	0.404	0.484	0.567
	TN	240	0.422	0.151	0.319	0.425	0.571
PAC-RSP02 / HRCA,2/2004	CA	663	0.861	0.105	0.806	0.875	0.933
	IL	461	0.843	0.119	0.786	0.865	0.926
	MO	218	0.801	0.128	0.716	0.815	0.889
	OH	598	0.797	0.113	0.736	0.810	0.857
	PA	570	0.847	0.103	0.790	0.862	0.917
	TN	240	0.800	0.119	0.736	0.826	0.875

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-BAL01 / HRCA,2/2004	CA	663	0.302	0.162	0.184	0.300	0.410
	IL	461	0.223	0.161	0.096	0.193	0.355
	MO	218	0.221	0.144	0.121	0.200	0.330
	OH	598	0.186	0.139	0.082	0.162	0.278
	PA	570	0.219	0.142	0.109	0.200	0.314
	TN	240	0.226	0.138	0.115	0.218	0.300
PAC-ROM01 / HRCA,2/2004	CA	663	0.519	0.151	0.419	0.517	0.593
	IL	461	0.529	0.156	0.432	0.532	0.643
	MO	218	0.506	0.145	0.435	0.514	0.619
	OH	598	0.471	0.180	0.364	0.487	0.564
	PA	570	0.544	0.165	0.438	0.555	0.619
	TN	240	0.517	0.153	0.418	0.513	0.584
PAC-ADL0X1 / Validation Report	CA	663	0.597	0.184	0.461	0.602	0.731
	IL	461	0.649	0.206	0.500	0.672	0.807
	MO	218	0.657	0.180	0.543	0.643	0.800
	OH	598	0.697	0.165	0.583	0.723	0.821
	PA	570	0.718	0.146	0.630	0.748	0.821
	TN	240	0.695	0.170	0.594	0.708	0.813
PAC-ADL0X2 / Covariate report	CA	663	0.586	0.178	0.460	0.585	0.731
	IL	461	0.657	0.197	0.528	0.674	0.807
	MO	218	0.664	0.175	0.555	0.653	0.800
	OH	598	0.698	0.158	0.598	0.716	0.821
	PA	570	0.722	0.137	0.636	0.747	0.821
	TN	240	0.671	0.166	0.561	0.681	0.813
PAC-CNT0X1 / Validation Report	CA	663	0.613	0.137	0.521	0.618	0.708
	IL	461	0.552	0.150	0.453	0.556	0.657
	MO	218	0.556	0.142	0.450	0.563	0.654
	OH	598	0.554	0.132	0.462	0.542	0.636
	PA	570	0.591	0.143	0.500	0.595	0.684
	TN	240	0.619	0.147	0.513	0.632	0.714
PAC-CNT0X2 / Covariate report	CA	663	0.590	0.126	0.508	0.597	0.708
	IL	461	0.574	0.144	0.482	0.581	0.657
	MO	218	0.588	0.137	0.504	0.605	0.654
	OH	598	0.521	0.131	0.430	0.515	0.636
	PA	570	0.548	0.128	0.464	0.546	0.684
	TN	240	0.592	0.145	0.498	0.600	0.714

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-RSP0X1 / Covariate Report	CA	663	0.206	0.103	0.135	0.194	0.286
	IL	461	0.199	0.116	0.116	0.182	0.272
	MO	218	0.239	0.120	0.153	0.225	0.316
	OH	598	0.236	0.105	0.163	0.227	0.318
	PA	570	0.180	0.097	0.110	0.169	0.250
	TN	240	0.239	0.115	0.158	0.231	0.321
PAC-RSP0X2 / Covariate Report	CA	663	0.213	0.104	0.137	0.201	0.286
	IL	461	0.208	0.118	0.124	0.196	0.272
	MO	218	0.243	0.124	0.160	0.231	0.316
	OH	598	0.236	0.102	0.160	0.224	0.318
	PA	570	0.182	0.093	0.117	0.167	0.250
	TN	240	0.252	0.116	0.172	0.240	0.321
PAC-WAL0X1 / Validation Report	CA	663	0.325	0.140	0.231	0.316	0.413
	IL	461	0.294	0.156	0.179	0.275	0.389
	MO	218	0.301	0.146	0.200	0.298	0.382
	OH	598	0.273	0.132	0.181	0.267	0.362
	PA	570	0.255	0.130	0.164	0.234	0.348
	TN	240	0.265	0.133	0.169	0.250	0.353
PAC-WAL0X2 / Covariate report	CA	663	0.331	0.137	0.243	0.330	0.413
	IL	461	0.278	0.146	0.171	0.266	0.389
	MO	218	0.284	0.142	0.185	0.284	0.382
	OH	598	0.280	0.132	0.186	0.274	0.362
	PA	570	0.262	0.125	0.173	0.252	0.348
	TN	240	0.281	0.132	0.184	0.268	0.353
PAC-WAL02 / HRCA,2/2004	CA	663	0.349	0.143	0.246	0.349	0.439
	IL	461	0.294	0.152	0.184	0.287	0.413
	MO	218	0.309	0.161	0.188	0.303	0.429
	OH	598	0.263	0.135	0.160	0.251	0.345
	PA	570	0.274	0.128	0.185	0.273	0.347
	TN	240	0.276	0.142	0.169	0.259	0.333
PAC-COG01 / Validation Report for Chronic (dropped)	CA	663	0.041	0.045	0.000	0.030	0.065
	IL	461	0.048	0.051	0.000	0.037	0.073
	MO	218	0.050	0.054	0.000	0.040	0.080
	OH	598	0.040	0.043	0.000	0.032	0.064
	PA	570	0.080	0.066	0.032	0.067	0.116
	TN	240	0.045	0.046	0.000	0.037	0.068

Table 1: Distribution of Post-Acute Quality Measures in Six Sampled Validation States[1]

QM Names/Source	State	Number of Facilities	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
PAC-CNT04 / Validation Report for Chronic	CA	663	0.238	0.101	0.170	0.229	0.303
	IL	461	0.189	0.101	0.117	0.182	0.250
	MO	218	0.219	0.106	0.150	0.200	0.287
	OH	598	0.253	0.099	0.177	0.250	0.313
	PA	570	0.243	0.108	0.167	0.235	0.314
	TN	240	0.229	0.111	0.154	0.217	0.308
PAC-RES01 / Currently reported (for Chronic) (dropped)	CA	663	0.087	0.083	0.026	0.065	0.130
	IL	461	0.026	0.045	0.000	0.000	0.036
	MO	218	0.034	0.043	0.000	0.020	0.050
	OH	598	0.046	0.062	0.000	0.027	0.067
	PA	570	0.031	0.047	0.000	0.000	0.046
	TN	240	0.066	0.067	0.015	0.047	0.103

[1] Analyses were based on all facilities in the six states.

Linking Quarterly QMs and Validation Database Data for Validation Analyses

In order to link chronologically each facility's QM measure to the validation items from the Validation Database, we chose the date of the facility walk-through[18] as a reference date. The following conventions were adopted:

- QMs for Q4 2001 were used for facilities with a walk-through date prior to January 14, 2002;
- QMs for Q1 2002 were used for facilities with a walk-through date between January 15, 2002 and April 14, 2002; and
- QMs for Q2 2002 were used for facilities with a walk -through date after April 15, 2002.

4.3 PAC Facility Selection

CMS asked the project team to validate proposed PAC QMs using data from two facility samples:

- Transitional care units (TCUs, originally used to validate PAC QMs in 2002); and

[18] "*Environmental Walk Through/ Resident Observation.* The aim of the Environmental Walk Through/ Resident Observation was to gain an overall understanding regarding whether the facility is "resident-centered", what the "feel" of the facility is, and what the nature of staff interactions with residents are (In the walk-through), measures were collected through assessment, surveillance, and observation of staff technique. The data collectors on site recorded their observations three times per day at approximately 10:00 a.m., lunchtime and 2:00 p.m. to obtain a comprehensive picture of the facility care environment." Morris, J.N. Validation of Long-Term Care and Post-Acute Quality Indicators. Final Draft Report (Version 2). September 27, 2002. Page 13.

- The full sample of nursing facilities. For PAC analyses, the full sample is defined as 196 facilities (facilities with one or more 5-day assessments in the data collection period for this study, from July 1, 2001 through June 30, 2002).

The reason for attempting to use the larger all-facility sample for validation testing was straightforward. It was thought that expanding the facility sample might increase the number of PAC QMs with statistically significant relationships to validation measures. However, adding more facilities could also create measurement problems. PAC QMs are meant to capture outcomes of care provided to short-stay residents who have potential for stabilization or rehabilitation. Yet there may be some facilities that do not provide "PAC-like" care to their short-stay residents. (For example, some might serve as way stations for residents who receive little rehabilitation, but move on to other settings after relatively brief stays). In the first Validation Report, TCUs were chosen to validate PAC QMs because it was presumed that their affiliation with hospitals meant that they would, for the most part, specialize in post-acute rehabilitative care. If the sample was expanded beyond TCUs, were there ways to infer from the available data which facilities were more likely than others to be providing PAC care to their short-stay residents? [19]

In a series of discussions, the team and CMS agreed on some "behavioral" dimensions of short-term nursing facility care that could be used to refine the sample. On the assumption that PAC care plans should aim to stabilize or rehabilitate, facilities that adhere to PAC-like care practices should not admit large numbers of short-stay residents with very little or no potential for stabilization or rehabilitation. A facility may be said to adhere to a PAC model if it does not admit high percentages of residents:

- with severe cognitive impairment;
- with severe late-loss ADL limitations;
- with high levels of clinical complexity; or
- in hospice care, who are end stage, or are comatose.

Using data from 5-day assessments in the Validation Sample of nursing facilities across the four quarters used in calculation of the QMs (from July 1, 2001 through June 30, 2002), the following measures that could capture these characteristics were computed separately for the TCU (N=60) and other facilities (N=136), as well as for all facilities combined:

1. Severe cognitive impairment: Facility percent of 5-day assessments with Cognitive Performance Scale >= 4

2. Severe late-loss ADL limitations: Facility percent of 5-day assessments with ADL score>= 16

3. High levels of clinical complexity: Facility percent of 5-day assessments with clinical complexity score = 3

4. Hospice: Facility percent of 5-day assessments with P1ao checked.

[19] It is important to note that, except for exclusions based on insufficient data or presences of certain conditions (end stage disease, for example), CMS does not distinguish among facilities in measuring PAC QMs. There is no federal standard for defining a "PAC facility." This project's attempt to refine the facility definition is designed to reduce measurement error in the validation analysis, not to suggest a facility taxonomy for QM computation and public reporting.

5. End stage: Facility percent of 5-day assessments with J5c checked.

6. Comatose: Facility percent of 5-day assessments with B1 = 1.

The first three of these measures are derived from the Resource Utilization Group-III (RUG-III) system, now used to adjust Medicare payments to nursing facilities. In particular, scales created from the RUG Nursing Home Case Mix Index (CMI) model include Clinically Complex, Late Loss ADL, and the Cognitive Performance Scale (CPS).

After inspecting the distributions, the team defined "outliers" for each measure as any value above the highest decile of the distribution across the entire sample of facilities (TCU and others). Outlier status meant that, on that measure, a facility was in the highest 10 percent of all facilities in severity or level of impairment at admission (5-day assessment). This led to the following list of threshold-defined outliers:

1. (CPS: Facility percent of 5-day assessments with CPS score >= 4) > 0.40

2. (ADL: Facility percent of 5-day assessments with ADL score >= 16) > 0.45

3. (Clinical complexity: Facility percent of 5-day assessments with clinical complexity score = 3) > 0.26

4. (Hospice: Facility percent of 5-day assessments with P1ao checked) > 0.01

5. (End stage: Facility percent of 5-day assessments with J5c checked) > 0.07

6. (Comatose: Facility percent of 5-day assessments with B1 = 1) > 0.01

To this list, the team added a seventh outlier criterion: the facility had fewer than 20 PAC 14-day assessments, over the four quarters selected for QM calculation.

Next, the team converted outlier status into categorical variables and computed "strings" of 0's and 1's in order to determine which facilities fell into multiple outlier categories. The results of this exercise were the following:

1. Of the 60 TCU facilities, 44 (73 percent) were not in any outlier category. Of the 136 non-TCUs, 78 (57 percent) were not in any outlier category. This was encouraging, because the team expected the TCU sample to conform more closely to the PAC model than the non-TCU sample.

2. Four facilities in each group had fewer than 20 14-day assessments.

3. In the TCU sample, 15 facilities were in only one outlier category. This meant that 59 of 60 had zero or one outlier characteristic. The one exception was an outlier on five counts (CPS, ADL, Clinical, Comatose and 14-day assessment).

4. In the non-TCU sample, 32 facilities were in only one outlier category, while 19 were in two, five were in three and two were in four.

It might be argued that one or more outlier criteria deserve higher weights in deciding which facilities to include or exclude. For example, being an outlier on ADLs might represent a more serious compromise of the PAC model than being a CPS outlier. However, the team used an equal weighting scheme, absent any general clinical guidelines that would support a more creative approach.

The following decision rules were adopted:

1. Exclude any facility that shows up in three or more outlier categories. This meant dropping one TCU facility and seven non-TCU facilities.

2. Exclude any facility with fewer than 20 14-day assessments. Three additional TCU and non-TCU facilities were dropped.

With these exclusions, the "refined full" sample for the validation analyses included 182 facilities:

- TCU - 56 facilities (down from 60).
- Non-TCU - 126 facilities (down from 136).

4.4 Validation Analyses

The primary goal of the validation analyses was to identify the PAC QMs that reflect the quality of post-acute care provided in the 209 nursing facilities that comprised our 2001-2002 multi-state validation sample. From the primary data collected in these sampled facilities (under the auspices of the MegaQI project), we selected a master list of validation variables that measure the care processes, management structure and general environment of a facility. We assume that these validation variables are on the causal-pathways between facility practice and the quality of care rendered in these nursing facilities. We hypothesized that there would be statistically significant associations between valid QMs and the validation variables, and that the direction of these associations would be compatible with assumed characteristics both of the validation measure and of the QM.

Selection of Validation Variables

The validation variables could be individual data elements collected in Administrative Survey (AS), Facility Walk Through (WAM) and Medicare Record Review (MRR) (referred to as items), or they could be any of 160+ validation scales derived from the items (referred to as scales) by the MegaQI Team for the validation analyses conducted (and initially reported) in 2002. By assumption, the validation variables were divided into two broad categories: preventive variables and responsive variables.

- Preventive variables capture policies or actions that facilities implement in advance, to minimize the emergence of problems.

- Responsive variables capture actions that facilities may use as they recognize that residents have ongoing or emerging problems.

This taxonomy creates an expected pattern of the direction of association. On the one hand, preventive variables should be negatively associated with QMs that measure deterioration or no improvement, but they should be positively associated with QMs that measure improvement or no impairment. On the other hand, responsive variables should be positively associated with QMs that measure deterioration or no improvement, but negatively associated with QMs that measure improvement or no impairment.

Selection of Items

For the validation analyses, we initially selected all the preventive items that had been used to validate either chronic or post-acute QMs in the MegaQI study. Overall, 42 preventive items were

selected from the Administrative Survey and Facility Walk Through[20]. We selected the preventive items for PAC QM candidates using the same selection method employed in the MegaQI project.

Responsive items were summary variables derived from the Medical Record Review (MRR), a structured medical record review. Using the MRR, nurse assessors reviewed medical charts of over 5,000 sampled residents for evidence of care processes. In the MRR, the same set of seventeen questions were asked repeatedly about 21 different areas of care, resulting in 357 data items. The 357 items are named with a letter indicating the area of care (a through u), followed by a number indicating the sequence number of the question (1 through 17). The 21 areas addressed in the MRR are:

a. Cognition
b. Communication
c. Delirium
d. Mood
e. Behavior
f. ADL improvement
g. ADL decline
h. Mobility
i. Falls
j. Antipsychotic drugs
k. Pain
l. Restraints
m. Feeding tubes
n. Nutrition
o. Catheter
p. Bladder incontinence
q. Bowel incontinence
r. Infections
s. Pressure scores
t. Burns
u. Little or no involvement in activities.

Data elements in the MRR were dichotomized variables and collected at the resident level. These variables were then summarized at the facility level, reflecting the percentages of sampled residents with a target event recorded in their medical charts in a facility. Take variable A_1 as example. The distribution of A_1 is:

Variable	n	Mean	Standard Deviation	Minimum	Maximum
A_1	209	0.516	0.316	0	1.00

The distribution of A_1 means the average percentage of sampled residents with a comprehensive assessment on cognitive impairment was 51 percent for the 209 nursing homes. In some nursing homes, none of the sampled residents had a comprehensive assessment, whereas in others, all had a comprehensive assessment.

[20] Source: Validation of Long-Term and Post-Acute Care Quality Indicators June 10, 2003, Appendix K. An excel file was created based on Appendix K to summarize the items used for QM validation. See H:\HSRE\AFTERSHOCK\PAC ANALYSES\items.xls.

In the MegaQI project, for a given QM, only responsive items addressing the same area of care were used for the QM validation analysis; for all areas of care, the same nine responsive items were selected. Accordingly, we selected nine corresponding items for each PAC QM for our analysis. However, there were three exceptions. For three QMs, shortness of breath (RSP02), balance (BAL01), and range of motion (ROM01), no corresponding responsive items were collected through the MRR. Therefore, no responsive items were tested for the three PAC QMs.

Selection of Scales

Through an intensive process of meeting and discussion, the MegaQI Team specified over 160 validation scales, derived from the individual items in the Validation Database. These scales were designed to represent the composite constructs of care processes or practice patterns of nursing homes. From this master list, we selected validation scales that were significantly associated with any of the chronic or PAC QMs in the hypothesized direction in the MegaQI study[21]. The selected scales could be dichotomized, categorical, counts (e.g., 10+), or continuous. Overall, 45 preventive scales (designated with "+" signs) and five responsive scales (designated with "-" signs) were selected for use in the bivariate analysis.

Bivariate Analyses

After selecting scales and validation items, we calculated Pearson correlation coefficients between all the PAC QMs and selected validation items or scales. We used this process to select explanatory variables for multivariate models, used in the next step to validate the PAC QMs. The selection criteria for the final regression model differed between preventive variables and responsive items. The selection criteria were:

- The bivariate correlation of the item and QM had to be statistically significant at $p < 0.10$.
- The sign of the correlation had to be consistent, both with the nature of the QM and with the classification of the item:
  - As noted earlier, QMs can be reflective of deterioration (call these Type A) or improvement (Type B).
  - Responsive items should have positive relationships with high levels of or increases in problems. (The bigger the problem, the more the facility needs to respond to it). So a responsive item should be positively correlated with Type A QMs, but negatively correlated with Type B QMs.
  - Preventive items should have a positive relationship to low levels of or decreases in problems. (If you prevent it, it won't happen or it will get better). So a preventive item should be negatively correlated with Type A QMs, but positively correlated with Type B QMs.
- The item or scale had to be significantly associated, and in the right direction, with at least two QMs.

[21] Source: Validation of Long-Term and Post-Acute Care Quality Indicators June 10, 2003, Appendix I. An excel file was created based on Appendix I to summarize all the validation scales tested in the validation of chronic or acute QMs. There are 160+ validation scales. To reduce the workload, only scales that are significantly associated with QMs in the hypothesized direction at least once were selected for PAC QM validation analysis. See H:\HSRE\AFTERSHOCK\PAC ANALYSES\val_scale_list.xls.

5.0 Findings

Bivariate analysis

The bivariate analyses were conducted for the TCU-only sample and the refined all-facility sample, respectively, refined as described in section 4.3. Pearson correlation coefficients were calculated for each pair of PAC QMs and selected validation variables.

We compared the bivariate results obtained from the TCU-only and refined all-facility samples. It is important to note that the number of facilities included in the bivariate analyses was larger for all facilities than for TCUs. The actual numbers of facilities used in the analysis depended on the QM and the effect of QM-specific exclusion criteria. For the TCU-only sample, the correlation coefficients were often calculated from 50 or more facilities; for the all-facility sample, the coefficients were generally calculated from 150 or more facilities. For a given PAC QM, there were often more variables that were significantly associated with the QM in hypothesized directions (referred to as significant and supportive variables) in TCUs than in all facilities, with the exception of functional QMs. For these measures, there was a fair to good match in the validation variables identified for a QM from the two samples: about half of variables identified from the TCU-only sample were also significantly associated with the QM in the refined all-facility sample. According to the variable selection criteria, only those variables that were significantly associated with at least two QMs in the hypothesized direction were included in the multivariable modeling. We identified 35 preventive and five responsive variables for the TCU-only sample, versus 54 preventive and eight responsive variables for the refined all-facility sample.

Upon this review of the bivariate results, the project team, with CMS, decided to use the refined "all-facility" sample of nursing facilities to report validation findings. This decision was based on the degree of similarity between the two samples in overall validation statistics. Additionally, the results from the refined "all facility" group can be generalized to a larger group of nursing facilities, both TCU and other, that admit and treat PAC residents.

Appendix 3 displays the significant and supportive associations in the all-facility sample by QM. Many of the items and scales that appeared in the earlier MegaQI validation models passed the selection criteria this time as well. Scales and items that met selection criteria for one QM in a related "family" (e.g., PAC RSP0X1 and PAC RSP0X2) generally, though not inevitably, appeared throughout the family.

Regression Results

Based on criteria listed in Section 4.4, we chose a set of validation variables for each PAC QM for multivariate modeling. Three multivariate models were estimated for each PAC QM. One included only preventive variables, one included only responsive variables, and the third included both preventive and responsive variables. The multiple R's from the multivariable regression models were the main indications for us to judge the validity of a PAC QM.

In Table 2 we summarize the results from the multivariate regression modeling. The rows of the table reference the individual QMs (listed in the same order as in the Technical Specification of PAC QMs in Appendix 1). In Table 2, we list QMs' sequence number, labels/names, their descriptions, and seven additional data elements. Columns four, five and six present the counts of significant, supportive validation variables for each QM, with separate counts for the number that fall under the preventive and responsive domains, and a final count of the total number of supportive validation elements for the indicator. Columns seven through nine provide the Multiple R correlation estimates

of the relationship between the pool of significant and supportive validation variables and the quality measure. The last column in the table, labeled "Degree of Validity", provides the final assessment of the confidence one can have in the quality indicator at the end of this validation process. There are three possible classifications:

- Level I, Highest Validity, represents those quality indicators with the strongest support.
- Level II, Moderate Validity, achieved lesser support but are still considered to be valid.
- Level III, Not Validated, represent measures that failed to be supported in this analysis. In their current form, there is insufficient reason to believe that they provide a reasonable facility estimate for the quality problems they seek to address.

Below is a summary of our findings.

- Of the 26 PAC QMs tested, 19 had Level I validity, one had Level II and six had Level III.

- Of the three PAC QMs currently reported, DEL0X1 had Level I validity, PAI0X1 had Level I validity and PRU0X2 had Level III validity. Our results were similar to the findings in the 2003 Validation Report of the MegaQI Project: in that report, the validity levels of these three QMs were Levels I, I and III, respectively

- Additionally, we compared our results to the validity levels of three QMs that were tested in all facilities in the MegaQI project, but were not public reported. The validity of CNT0X1, RSP0X1 and WAL0X1 were all Level I.[22]

- The validity of the QMs addressing the same care areas or conditions was often similar. For example, although ADL QMs have different denominators, numerators and exclusions, all achieved Level I validity.

- QMs with the same numerators and denominators often performed similarly in the validation analyses.

.

[22] For details see Validation Report, Appendix M. Our results were comparable to MegaQI validation results on all facilities without FAP. The QMs were named cnt0x, rsp0x and wal0x.

Table 2: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
1	PAC-DEL0X1/ Currently reported	Percent of short-stay residents with delirium	5	3	8	0.53	0.47	0.66	I
2	PAC-DELOX2/ Covariate report	Percent of short-stay residents with delirium	5	3	8	0.53	0.46	0.66	I
3	PAC-PAIOX1 / Currently reported	Percent of short-stay residents who had moderate to severe pain	6	0	6	0.53	---	0.53	I
4	PAC-PAIOX2 / Covariate report	Percent of short-stay residents with pain	5	0	5	0.47	---	0.47	I
5	PAC-PRU0X1	Percent of short-stay residents whose pressure sores have not gotten better.	1	1	2	0.23	0.28	0.38	III
6	PAC-PRU0X2/ Currently reported	Percent of short-stay residents whose pressure sores have not gotten better.	0	0	0	---	---	---	III
7	PAC-PRU0X3 / Covariate report	Percent of short-stay residents whose pressure sores have not gotten better	1	1	2	0.25	0.24	0.36	III
8	PAC-ADL04 / HRCA,	Percent of residents with improving level	8	0	8	0.79	---	0.79	I

Table 2: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
	2/2004	of ADL functioning							
9	PAC-ADL05 / HRCA, 2/2004	Percent of residents who improve status on mid-loss ADL functioning (transfer, locomotion) or remain completely independent in mid-loss ADLs	11	0	11	0.80	---	0.80	I
10	PAC-ADL06 / HRCA,2/2004	Percent of residents who improve status on early-loss ADL functioning (dressing and personal hygiene) or remain completely independent in early-loss ADLs (ELADL).	7	0	7	0.68	---	0.68	I
11	PAC-MOD04/ HRCA,2/2004	Percent of residents who improve their mood or remain free from symptoms of depression (based on MDS Depression Rating Scale)	0	1	0	---	0.23	---	III
12	PAC-CAT03/ HRCA,2/2004	Percent of residents who do not have a catheter at 14-day assessment	0	0	0	---	---	---	III

Table 2: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
13	PAC-PAI02/ HRCA,2/2004	Percent of residents who improve their pain status or remain free from pain, (based on the MDS Pain Scale)	9	1	10	0.55	0.25	0.58	I
14	PAC-RSP02/ HRCA,2/2004	Percent of residents who do not have shortness of breath	7	0	7	0.65	---	0.65	I
15	PAC-BAL01/ HRCA,2/2004	Percent of residents who improve their balance function or remain free from impairment in balance function between 5 and 14-day assessment	8	0	8	0.75	---	0.75	I
16	PAC-ROM01/ HRCA,2/2004	Percent of residents who improve their range of motion or remain free from impairment in ROM between 5 and 14-day assessment	0	0	0	---	---	---	III
17	PAC-ADL0X1/ Validation Report	Percent of residents who have not improved since admission	14	0	14	0.81	---	0.81	I
18	PAC-ADL0X2/ Covariate report	Percent of short-stay residents who have not improved since admission	11	0	11	0.70	---	0.70	I

Table 2: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
19	PAC-CNT0X1/ Validation Report	Failure to Improve Bladder Incontinence	5	1	6	0.47	0.24	0.47	I
20	PAC-CNT0X2/ Covariate report	Failure to Improve Bladder Incontinence	5	0	5	0.49	---	0.49	I
21	PAC-RSP0X1/ Covariate Report	Percent of short-stay residents who have developed a respiratory infection or have not gotten better	5	0	5	0.57	---	0.57	I
22	PAC-RSP0X2/ Covariate Report	Percent of short-stay residents who have developed a respiratory infection or have not gotten better	4	1	5	0.48	0.24	0.52	I
23	PAC-WAL0X1 / Validation Report	Percent of residents who walk as well or better on day 14 as on day 5 of their stay.	8	0	8	0.72	---	0.72	I
24	PAC-WAL0X2 / Covariate report	Percent of short-stay residents who walk as well or better on day 14 as on day 5 of their stay	9	0	9	0.71	---	0.71	I

Table 2: Summary Measures of Quality Indicatory Validity[1]

Order in Appendix 1	Label	Quality Indicator	Count of Significant Preventive Data Elements[2]	Count of Significant Responsive/ Reactive Data Elements[2]	Total Count of Significant Data Elements[2]	Multiple R for Preventive Elements	Multiple R for Responsive Elements	Multiple R for All Elements	Degree of Validity I TOP II MID III NOT Valid[3]
25	PAC-WAL02/ HRCA,2/2004	Percent of residents with improving level of locomotion functioning or who remained completely independent in locomotion functioning	5	0	5	0.66	---	0.66	I
26	PAC-CNT04 / Validation Report for Chronic	Prevalence of urinary tract infections	1	1	2	0.27	0.30	0.43	II

Notes:

1 The analyses were based on data from "TCU-only" sample.

2 An alpha significance level for the correlation between the validation element and the quality indicator of .10 or lower. Note that these counts refer to the count of elements entered into the multivariate models.

3 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

--- Indicates that statistics could not be generated due to lack of significant data elements.

6.0 Discussion

As displayed in Table 2, the majority (77 percent) of these "new" post-acute care quality measures appear to be valid, at either high (Level I) or mid (Level II) validity levels. This finding will likely be met with enthusiasm by regulators, policy-makers, consumers and advocates, as well as by the long-term care industry. There is a desire among these various stakeholders to better capture post-acute care quality, and such a large set of new indicators provides additional means with which to publicly report and assess the quality of care rendered. A large set of measures across multiple domains of care provides multiple views of care quality, which will be beneficial to providers who seek to improve their quality of care processes. Given the largess of the set, however, there will be a need to select among and prioritize the valid indicators for public reporting and other uses. In this discussion we attempt to provide CMS and others with assistance in interpreting and "digesting" the nuances, strengths and potential limitations of the individual and collective measures evaluated here. Table 3 categorizes each QM by the characteristics described below.

QMs by domain. There are several domains of care and treatment represented by this set of post-acute care quality measures, including: clinical complexity; functional status; mobility (which may also be categorized as an aspect of physical functioning); and mood/behavior. The care areas, level of functioning or health status represented within these domains are broad, and exceed the number of PAC QMs validated through the earlier MegaQI study. In the domain of clinical complexity, we present here four separate clinical concepts for which at least one QM representing that concept has Level I or Level II validity: delirium; pain; bladder functioning; and respiratory functioning.

We have also expanded the breadth of the functional status domain, by presenting not only improvement measures, which we believe can encourage positive practices among providers, but measures that address aspects of functioning not yet addressed and very relevant to the post-acute care population. For example,

- Two new measures presented here capture improvements in "early-loss" and "mid-loss" (transfer and locomotion) ADLs. Early loss ADLs include the tasks of dressing and personal hygiene, that represent higher levels of physical functioning.[23] These new functional measures attempt to assess the ability of nursing facilities to assist short-stay residents to improve (or in many cases, regain) functioning in these functional areas that may have been lost or weakened by surgery, a lengthy illness or infection in hospital, stroke or other impediment to their usual ability to perform these various ADLs.

- Three other functional quality measures (one of which was reported in the original MegaQI validation report) tested in this study are derived from the ADL Long Form summary scale, which has a range from zero to 28 and includes all seven MDS ADL self-performance items, all of which are believed to be associated with early, middle and late loss functioning.

QMs by prevalence vs. incidence, and by improvement vs. decline. Two of the three publicly reported PAC QMs are all cross-sectional QMs (prevalence), though the original MegaQI validation study did find four valid PAC incidence (change in status) measures. In our current analyses, we validated both cross-sectional and incidence measures. Nineteen of the 26 PAC measures are incidence measures; 14 incidence measures total were found to have Level I or Level II validity.

[23] Morris JN, Fries, BE, Morris, SA. Scaling ADLs within the MDS. J. Gerontol. A Biol Sci Med Sci. 1999 Nov; 54(11):M546-53.

Change in status QMs, especially those that capture improvement in function and clinical condition, are important for monitoring post-acute care, as the primary focus of post-acute care is on the restoration of residents' function. Aggregated statistics on changes in residents' clinical condition and functioning at the facility-level are important measures of how well facilities perform in this restorative aspect of care. Also, QMs that measure improvement and that measure deterioration may be conceptually different, and may reflect different care practices. For example, higher rates for "improvement" QMs may result from more aggressive clinical interventions, whereas lower rates for "deterioration" QMs may be the result of better prevention strategies. Accordingly, we may need to develop and test a family of QMs that address different aspects in the same domain or care area. In addition, we favor providing consumers with data on both bad as well as good facility performance to better inform their healthcare decisions. Further development of both types of QMs may enable us to provide a more complete picture of the quality of care of a facility, and allow CMS to select relevant QMs from a larger pool of candidates.

Table 3 categorizes PAC QM candidates that address both potential deficits and positive achievements in post-acute care. Of the 26 measures evaluated, there are 10 QMs that describe facilities' performance in improving/maintaining residents' function and clinical or emotional status, and nine that measure deterioration or lack of improvement.

QMs by adjustment model. In this set of 26 measures, there are five concepts or care areas represented by multiple models or forms of the QM. For example, there are three pain measures, all found to have high (Level I) validity. Two of the pain measures are prevalence measures, and one is a change in status measure (residents improve pain status or remain free from pain). Of the two prevalence measures, one (PAI0X1) is unadjusted, while the other (PAI0X2) is risk-adjusted with resident-level covariates (CPS, Hip Fracture and Alzheimer's/Other Dementia). The pain incidence measure (PAI02) captures the proportion of residents with improvement in their Pain Scale Score (or who remain free from pain) between admission and the 14-day MDS assessment. This measure has resident-level covariates (CPS and hip fracture).

As has long been the stance of this project team, we favor risk-adjusted measures over unadjusted, in most cases, and hope that CMS will consider some of the validated measures adjusted with RUG-based covariates, since they should better capture casemix-related risk for those outcomes they are associated with.

Table 3 categorizes each QM by the factors discussed above (e.g., quality domain, prevalence vs. incidence, risk-adjustment). Section 7 provides some further discussion about how CMS might approach the task of determining which, among this set of validated post-acute care quality measures, are appropriate for public reporting.

Table 3: QM Characteristics

Domain	Quality Measure	Concept Represented	Type of Measure	Risk Adjustment
Clinical Complexity	PAC-DEL0X1	Delirium	Prevalence	Prior residential history
	PAC-DEL0X2	Delirium	Prevalence	None
	PAC-PAI0X1	Pain	Prevalence	None
	PAC-PAI0X2	Pain	Prevalence	CPS, Hip Fracture, Alzheimer's/Other Dementia)
	PAC-PAI02	Pain	Incidence, Improvement	CPS, Hip Fracture
	PAC-PRU0X1	Pressure Sores	Incidence, Decline	None
	PAC-PRU0X2	Pressure Sores	Incidence, Decline	Unresolved pressure sore, bed mobility, bowel incontinence, DM or PVD, low BMI
	PAC-PRU0X 3	Pressure Sores	Incidence, Decline	R_CLN, R_CMI, R_ADL
	PAC-CAT03	Bladder Function	Prevalence	R_ADL, R_CLN
	PAC-CNT0X1	Bladder Function	Incidence, Decline	None
	PAC-CNT0X2	Bladder Function	Incidence, Decline	R_CMI, R_ADL
	PAC-CNT04	Bladder Function	Prevalence	None
	PAC-RSP0X1	Respiratory Function	Incidence, Decline	Asthma, Emphysema/COPD
	PAC-RSP0X2	Respiratory Function	Incidence, Decline	R_CLN
	PAC-RSP02	Respiratory Function	Prevalence	Emphysema/COPD
Functional Status	PAC-ADL04	ADLs	Incidence, Improvement	CPS
	PAC-ADL0X1	ADLs	Incidence, Decline	No prior residential history
	PAC-ADL0X2	ADLs	Incidence, Improvement	CPS
	PAC-ADL05	ADLs	Incidence, Improvement	CPS, R_ADL
	PAC-ADL06	ADLs	Incidence, Improvement	CPS, R_ADL
	PAC-BAL01	Balance	Incidence, Improvement	CPS, R_ADL
	PAC-ROM01	Range of Motion	Incidence, Improvement	R_ADL, hip fracture
Mobility	PAC-WAL0X1	Walking	Incidence, Improvement	None

Table 3: QM Characteristics

Domain	Quality Measure	Concept Represented	Type of Measure	Risk Adjustment
	PAC-WAL0X2	Walking	Incidence, Improvement	R_ADL, R_CMI, CPS
	PAC-WAL02	Locomotion	Incidence, Improvement	CPS, R_ADL
Mood	PAC-MOD04	Depression	Incidence, Improvement	None

7.0 Conclusions and Next Steps

Together with colleagues from HRCA, Abt Associates accomplished the two principal objectives of this project: 1) to acquire, test and deliver to CMS a documented copy of the Validation Database, and 2) to identify and validate additional candidate PAC QMs. Clearly, the second objective depended heavily on success in attaining the first. Although it would have been possible to implement a new approach to validating the new set of PAC QMs, one that did not depend on replicating the earlier approach, it was decided that consistency in methodology would be important to securing acceptance of our findings. Therefore, it was critical that the team fully understand the processes used to validate QMs in the 2003 Validation Report. We demonstrated this understanding by replicating statistics reported in the 2003 Validation Report.

We also made substantial progress in expanding the number of valid PAC QMs.

- We were able to demonstrate high or moderate levels of validity for 20 additional PAC QM measures.
- Our findings did not substantially change the validity determinations for the three PAC QMs that are now publicly reported.
- Compared to the 2002 analyses, which validated PAC QMs on TCUs only, we used a more broadly representative sample of nursing facilities.

As depicted in Tables 2 and 3, there are multiple, valid measures of post-acute care quality presented here from which to choose for public reporting (or other purposes). Represented in this measure set are several desirable features for quality measurement, such as the use of resident-level risk adjustment, and more incidence measures depicting improvement in status than previously available for post-acute care.

Though we hesitate to recommend any one measure over another, we do encourage CMS to consider the potential benefit of having multiple measures of a given clinical or functional concept, in order to best capture the multidimensionality of nursing facility care quality. In this vein, it may be beneficial to publicly report an incidence and prevalence pain measure, for example, or an improvement and decline measure for the sum of ADLs presented in the ADL Long Form summary scale (e.g., ADL04 and ADL0X2). As stated previously, we do recommend those QMs that are risk-adjusted, over others that are not.

At the end of the MegaQI project, we identified additional analyses that CMS might consider for extending or refining the current QM system. Here, we note additional analyses that might be undertaken to expand upon and contribute to this body of work.

- QMs might be developed and tested for "special populations" admitted for post-acute care, such as dementia residents or residents with special mental health needs. Presently, QMs apply to broadly defined resident groups. Some include all residents in a facility. Some measure problem prevalence separately for high- and low-risk cases. Some were presumed to be valid only after exclusions of certain special groups (for example, hospice residents and residents with end stage diseases). This practice raises a question of how to detect variations in the quality of care provided to "excluded" groups. Hypothetically, new QMs might better capture dimensions of the quality of care provided to these excluded groups.

- Sensitivity tests for validity might be conducted. The MegaQI Team assigned levels of validity based on statistical measures of association. In some cases, these measures of association varied narrowly around thresholds for accepting or rejecting a QM. It was easy to see that a change of one or two percentage points in a coefficient of determination could have reversed a decision about the validity of a QM. Further exploration of the validity of the current set of measures, using differing methods, may yield useful information regarding the relationship of the measures to facility process measures of quality.

- The use of facility-level risk adjustors continues to show promise for producing QMs with high discriminatory power. Further work performed by HRCA on a direct adjustment method for capturing facility casemix also shows promise. These issues warrant further exploration and eventual adoption for public reporting.

- One large task of the MegaQI contract, that of establishing benchmarks for ranking providers on the quality of care they provide, was curtailed due to the more pressing needs of public reporting. Work should proceed on the establishment of benchmarks and on the appropriate manner in which to display QM data to enable clear consumer understanding.

- The current Administration has shown interest in "pay for performance" systems, and in their implementation, possibly on a demonstration basis, in nursing homes, home health agencies and physicians offices (hospital pay for performance demonstrations and voluntary programs are now underway). In advance of a mandated demonstration, CMS might undertake an assessment of what changes, if any, in the current system of quality measures might be appropriate to support nursing home pay for performance (for example, could the current QMs provide the basis for a scoring system similar to the one currently in place for the Premier hospital pay for performance demonstration?).