Published in final edited form as:

Med Care. 2008 May ; 46(5): 532–541. doi:10.1097/MLR.0b013e31816099c5.

Does risk adjustment of the CMS quality measures for nursing homes matter?

Dana B. Mukamel, Ph.D.¹, Laurent G. Glance, M.D.², Yue Li, Ph.D.³, David L. Weimer, Ph.D.⁴, William D. Spector, Ph.D.⁵, Jacqueline S. Zinn, Ph.D.⁶, and Laura Mosqueda⁷

Dana B. Mukamel: dmukamel@uci.edu; Laurent G. Glance: laurent_glance@urmc.rochester.edu; Yue Li: yueli@buffalo.edu; David L. Weimer: weimer@lafollette.wisc.edu; William D. Spector: WSpector@AHRQ.GOV; Jacqueline S. Zinn: Jacqueline.Zinn@temple.edu; Laura Mosqueda: mosqueda@uci.edu

¹ Professor & Senior Fellow; University of California, Irvine, Center for Health Policy Research, 111 Academy, Suite 220 Irvine, CA 92697-5800; tel.: 949-824-8873, fax: 949-824-3388

² University of Rochester, Anesthesiology M&D, Associate Professor, 601 Elmwood Ave. Box 604, Rochester, NY 14642, Tel:585-275-4187, Fax (585) 244-7271

³ Assistant Professor, ECMC-Clinical Center, Room CC163, State University of New York at Buffalo, 462 Grider Street Buffalo, NY 14215 Tel: Phone: (716) 898-5175, Fax: (716) 898-3536

⁴ Professor, University of Wisconsin – Madison, LaFollette School of Public Affairs, 1225 Observatory Drive, Madison, WI 53706; tel.: 608-262-5713, fax: 608-265-3233

⁵ Senior Social Scientist, Agency for Healthcare Research & Quality; 540 Gaither Road, 5th Floor Rockville, MD 20852; tel.: 301-427-1446, fax: 301-427-1430

⁶ Professor, Temple University, 413 Ritter Annex Philadelphia, PA 19122-6091; tel.: 215-848-0796, fax: 215-848-4712

⁷ Professor; Family Medicine and Program in Geriatrics, University of California Irvine Medical Center, Building 200, Suite 835, Route 81; Tel: (714) 456-5530, Fax: (714) 456-7933

Abstract

Background—CMS publishes a report card for nursing homes with 19 clinical quality measures (QMs). These measures include minimal risk adjustment.

Objectives—To develop quality measures with more extensive risk adjustment and to investigate the impact on quality rankings.

Research Design—Retrospective analysis of individual level data reported in the Minimum Data Set (MDS). Random effect logistic models were used to estimate risk adjustment models for 5 outcomes: pressure ulcers for high and low risk patients, physical restraints, and pain for long- and short-stay patients. These models were used to create 5 quality measures with extended risk adjustment, EQMs. The EQMs were compared to the corresponding QMs.

Subjects—All (17,469) nursing homes that reported MDS data in the period 2001-2005, and their 9.6 million residents.

Measures—QMs were compared with EQMs for all nursing homes in terms of agreement on outlier identification: Kappa, false positive and false negative error rates.

Results—Kappa values ranged from 0.63 to 0.90. False positive and negative error rates ranged from 8% to 37%. Agreement between QMs and EQMs was better on high quality rather than on low quality.

Conclusions—More extensive risk adjustment changes quality ranking of nursing homes and should be considered as potential improvement to the current QMs. Other methodological issues

related to construction of the QMs should also be investigated to determine if they are important in the context of nursing home care.

Keywords

nursing homes; quality of care; risk adjusted outcome measures of quality; quality report cards; Nursing Home Compare

Introduction

Quality of care in nursing homes has and continues to be of concern to patients, their families, professionals and policy makers.¹⁻³ The Centers for Medicare & Medicaid Services (CMS), as part of their quality initiative, have begun publishing a quality report card that currently provides information about 19 clinical Quality Measures (QMs).⁴ The information is intended to inform patients' and families' choices of nursing homes, state regulators, and quality improvement efforts by nursing homes.^{5, 6}

The CMS report card, available on the web at <http://www.medicare.gov/NHCompare> and entitled Nursing Home Compare, was first published in November 2002. It included 10 QMs capturing the experience of both long-stay (chronic) and short-stay (post acute) residents. Since then, CMS has been improving the measures, deleting those that were judged to be inadequate and adding new measures. The current list of 19 measures is shown in table 1.

The literature on measuring quality based on health outcomes has long recognized that raw incidence or prevalence rates of adverse outcomes cannot be interpreted as measures of quality.⁷ Because health outcomes depend not only on the quality of care patients receive but also on patients' risk factors, outcome based measures of quality must account for differences in the mix of patients' risks across providers. Otherwise, these measures are likely to be biased, typically underestimating the quality of those treating the sickest patients. Previous studies have demonstrated that different risk adjustments methods lead to different quality rankings of hospitals.⁸⁻¹⁰ A study of nursing homes' quality demonstrated that outlier designation and ranking of facilities based on raw outcome rates differed as the degree of risk adjustment was increased, from only age adjustment, to age and a case mix index (RUGs II) adjustment, to a more comprehensive risk adjustment that included several factors capturing physical and mental health.¹¹ Berlowitz et al.¹² and Rosen et al.¹³ had similar findings for Veteran Affairs (VA) nursing homes. While there are other technical issues related to the construction of quality measures based on patient health outcomes,¹⁴⁻¹⁷ risk adjustment is one of the most critical, and is the focus of the work we present here.

The method used by CMS to calculate the majority of the nursing homes' QMs includes limited risk adjustment, operationalized mostly through exclusion rules (see table 1). The method consists of two steps: 1) It defines the relevant population (the denominator) based on a set of outcome specific exclusion criteria. For example the measure based on the rate of residents whose ADL limitations increased excludes those who are in hospice care or are terminally ill. 2) It calculates the QM as the percent of the population identified in step (1) that experienced the outcome. Thus the risk adjustment is introduced mostly via the exclusion rules which create more homogeneous comparison groups. Five of the 19 QMs also have an additional limited covariate risk adjustment (see table 1). This method differs from methods used in other report cards, including the New York State (NYS) Cardiac Surgery Report which is considered to be the gold standard among report cards.¹⁸ The latter adjusts for risk using multivariate regression techniques. Regression based methods allow for more flexibility in the risk adjustment and can accommodate a larger number of risk factors.

The decision by CMS to report QMs that have only a minimal risk adjustment followed a series of studies that examined measures with varying degrees of risk adjustment, ranging from measures similar to those finally adopted, based mostly on exclusion criteria^{19, 20} to measures with extensive risk adjustment based on multivariate regressions,^{21, 22} similar to the method used in the NYS Cardiac Surgery Reports. The CMS decision was based on recommendations by the National Quality Forum,^{23, 24} a private non-profit organization that engaged in a political consensus building process and reflects the views of a diverse group of stakeholders.

In this paper we examine the impact of further risk adjustment of several of the nursing homes' QMs by developing measures with a larger number of risk factors and by employing multivariate risk adjustment techniques. We compare facility designation based on the CMS QMs with those based on the enhanced quality measures (EQMs). We do not view the EQMs that we developed as *the* gold standard, but rather aim to demonstrate that the extent of the risk adjustment can substantially impact quality rankings and should, therefore, be considered as a potential improvement to the QMs.

Considerations in Choosing Risk Factors

Several issues should be considered when determining the extent and type of risk adjustment, including: bias due to incomplete risk adjustment, "giving credit" for poor quality, and usefulness of the quality measures for quality improvement programs.

The issue of bias due to incomplete risk adjustment arises because health outcomes are not only a function of quality of care, but also depend on the patient's health status. Unbiased quality measures will account for the frailty and severity of illness of the residents in each nursing home, thus holding the facility harmless for those patient characteristics that it cannot control, such as age and gender. To minimize the potential bias, the list of risk factors should be as comprehensive as possible, including all characteristics deemed to be clinically associated with the outcome. Omission of important patient risk factors may result in erroneous measures of quality, incorrect ranking of providers, and inappropriate labeling of providers as high or low quality outliers.

Some risk factors are, however, also outcomes of care. They depend both on the clinical condition of the individual and the care the individual received. For example, ADL limitations are risk factors for pressure ulcers, but they themselves are health outcomes that depend on quality of care. Excluding ADL limitations from the risk adjustment for pressure ulcers introduces bias in the pressure ulcers measure. Including them, however, "gives credit" to the poor care which led to the poor ADL outcomes, i.e. a facility with poor ADL care will have a better pressure ulcer quality measure if ADLs are included as risk factors because it has high values for these risk factors, everything else being equal. One might, therefore, be concerned that including ADLs as risk factors creates a perverse incentive to lower the quality of ADL care. Incorporating a separate quality measure based on ADLs in the report card guards against such perverse incentives. As long as those risk factors that are also health outcomes are included in the set of published measures, as is the case with ADLs, facilities will be penalized for poor ADL quality, and there is no need to penalize them again, indirectly, through the pressure ulcers measure.

Another important consideration is the usefulness of the quality measure to consumers and providers. The simplicity of the CMS approach to risk adjustment, based on exclusions from the denominator and creation of patients groups of different risk levels (e.g. patients at high and at low risk for pressure sores) makes the QMs easy to understand for both professionals and lay persons. This approach is intuitive because it allows one to imagine the group of patients who are defined in the denominator for the QM. It is thus more conducive to targeting quality improvement efforts by allowing identification of groups of patients whose outcomes, as

measured by the QMs, are poor. Multivariate regression based risk adjustment is much more complex and does not lend itself to this intuitive identification of patient groups whose care is poor.

This advantage has to be weighed against the disadvantage of the exclusion/stratification approach adopted by CMS, a disadvantage that arises due to the limited number of risk factors that can be accommodated. The limit on the number of risk factors is due to: 1) sample size – comparison groups become smaller the more restrictively each group is defined; and 2) complexity of exposition – the number of QMs increases with the number of risk factors and the number of required strata, increasing the number of comparison groups to a point that may become overwhelming for the users of the information. For example, making oneself understood is an important risk factor for pain. This variable has 3 levels – always, sometimes, and rarely. Gender is also an important risk factor. The CMS approach of creating homogenous groups would require 6 different groups for the pain measure, just to account for these two types of risk factors. It would require many more groups to account for all necessary risk factors.

The extensive risk adjustment, regression based approach has an additional advantage. The regression models, because they explicitly model the relationships between health outcomes and a comprehensive set of patient risks, reveal associations that can be useful in guiding quality improvement efforts, targeting individuals with specific risks. For example, the models predicting pressure ulcers show that loss of voluntary movement is an important predictor. This could guide development of interventions targeting those with loss of voluntary movement.

Methods

Data and sample

We used national MDS (Minimum Data Set) data for the period 2001-2005. The MDS includes individual level assessments for all residents in nursing homes that participate in Medicare or Medicaid. The data, collected at admission and at specific intervals post admission, include information about the person's socio-demographics, physical and mental health status, and specific treatments. These data have been collected by all nursing homes and submitted to CMS since 1998 and are used to calculate Medicare payment rates for nursing homes (based on RUGs III) and the QMs for all residents.²⁵ Many of the specific data elements have been shown to be of high reliability with intraclass correlations of 0.6 or higher²⁶ and Kappas for interrater reliability exceeding 0.7.^{27, 28} CMS conducts annual training to enhance data quality.²⁵

The five years of data we used included information for 9.6 million residents in 17,469 facilities.

Construction of CMS quality measures

Using the CMS algorithm²⁹ we calculated five QMs: percent of high risk residents with pressure ulcers; percent of low risk residents with pressure ulcers; percent of residents who are physically restrained, percent of long-stay residents with pain; and percent of short-stay residents with pain. The QMs were calculated for each quarter during 2001 through 2005, based on the Prospective Payment System (PPS), quarterly and annual MDS assessments available for the quarter.

Construction of enhanced quality measures

We developed the enhanced quality measures (EQMs) as follows:

1. We defined the outcome in the same fashion as the CMS method, applying the same inclusion and exclusion criteria.²⁹ (In essence, our outcome equals the CMS QM at

the individual level.) All outcomes were defined as dichotomous variables, set to 1 if the resident experienced the outcome, 0 otherwise.

2. We identified additional risk factors (beyond those included in the CMS QMs) from the relevant literature. We then included all those that were available in the PPS, quarterly and annual MDS assessments and that were judged by a geriatrician, who is familiar with the nursing home population, to potentially influence the outcome. Risk factors were identified separately for each outcome.
3. We then estimated hierarchical random effect logistic models to identify the association of each risk factor with the relevant outcome. These models were based on individual level outcomes and risk factors, allowing for a random facility intercept to account for the potentially non-random clustering of residents within facilities, as follows:

$$\text{Ln} \frac{P_{i,j}}{1 - P_{i,j}} = \alpha + \sum_k \beta_k X_{i,j,k} + \varepsilon_{i,j} + v_j \quad (\text{i})$$

where $p_{i,j}$ is the probability that resident i in facility j will experience the outcome, $X_{i,j,k}$ are the k risk factors for the resident, v_j is a facility error component and $\varepsilon_{i,j}$ is the resident error component.

Model specification was guided by inspection of the C statistic, which measures the discrimination of the model, and the Hosmer-Lemeshow statistic, which measures the calibration of the model.³⁰ Based on these criteria we determined that model fit improves when the risk models are stratified by age.³¹ We, therefore, estimated separate models for each of four age groups: age less than or equal to 64, 65-74, 75-84 and 85+. This is equivalent to interacting age with all other risk factors. Because of the large sample size there is no efficiency advantage in pooling the data.

All risk factors were included in the initial models and those not significant at the 0.25 level were excluded from the final models.

4. Because of the very large size of our dataset the traditional approach of randomly splitting the data into development and validation data sets³¹ was not feasible. Instead, we created multiple random samples of about 100,000 observations in each, and repeated the estimation for each sample. We calculated the average of each estimated coefficient across all samples and are presenting the estimates averaged over 200 iterations. We determined the stopping point of this process at 200 iterations by comparing the change in the average coefficients in increments of 50 iterations, i.e. we compared the average coefficient for the first 50 iterations to the average of the first 100, the first 100 to the first 150, and the first 150 to the average of 200 iterations. The change in the average coefficients between the 150 to 200 iterations was less than 2% for all, indicating that the estimates are converging to their true means within an acceptable margin of error.
5. The average coefficients (based on 200 iterations) were then used to calculate the expected probability of the outcome for each resident, $E_{i,j}$, conditional on his or her specific risk factors, as follows:

$$E_{i,j} = \frac{e^{\hat{\alpha} + \sum_k \hat{\beta}_k X_{i,j,k}}}{1 + e^{\hat{\alpha} + \sum_k \hat{\beta}_k X_{i,j,k}}} \quad (\text{ii})$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the average estimated coefficients from the regression models given by equation (i).

6. The expected and observed facility rates were calculated as the average of the individual expected probability and the average of the observed outcomes respectively.
7. The EQM was calculated as the ratio of the observed to expected rate.^{7, 15, 32}

Comparison of QMs and EQMs

We calculated QMs and EQMs for all facilities for the 4th quarter of each calendar year. Per CMS's method, we excluded facilities that had too few residents who qualify to be included in the denominator – fewer than 30 for the long-stay QMs and fewer than 20 for the short-stay QMs. We then compared the classification of nursing homes by the QMs and the EQMs, separately for each quarter and for each QM/EQM pair.

We compared the assignment of nursing homes to the tails of the distribution by the QMs and the EQMs. We examined outliers defined for three different cutoffs: 5, 10 and 20 percents, separately for the top and the bottom of the distribution. These are cutoffs that identify facilities with extreme performance and thus might be used in quality report cards or pay for performance systems.

To assess the agreement between the EQMs and the QMs on outlier designation we repeated the following analyses for the high quality and the low quality outliers separately.

We calculated the Kappa statistic, which is a measure of overall agreement in classification of outliers, above and beyond the agreement expected by chance alone. The Kappa ranges between 0 and 1, with 0 indicating no agreement beyond chance and 1 indicating perfect agreement. Kappa values between 0 and 1 indicate the proportion of cases on which the two measures agree, accounting for chance agreement.³³

We also calculated two measures of errors: 1) the false negative error rate, defined as the proportion of outliers identified by the EQM but not by the QM; and 2) the false positive error rate defined as the proportion of outliers identified by the QMs but not by the EQMs.³⁴ We calculated the false negative and false positive rate assuming that the EQMs are more accurate measures of quality than the QMs, and hence treat the EQMs as the “gold standard”.

Results

Table 2 shows the distribution of the five QMs for the 4th quarter of 2005. The number of facilities for which we report these data range from 6,732 to 13,639 because not all QMs are reported for all facilities due to small samples (less than 30 for the long-stay measures and less than 20 for the short stay-measures). The variation in the QMs is quite substantial, with coefficients of variation ranging from about 60% to almost 120%.

Table 3 provides descriptive statistics for all the variables used in the analyses. The table shows data separately for the long- and the short-stay measures because the populations for which these measures are calculated are different.

Table 4 shows the estimated coefficients for each of the risk adjustment models. The numbers shown are the averages based on 200 iterations of random samples with approximately 100,000 observations per iteration. Most risk factors were significantly associated with the outcomes. In a few cases, the p values exceeded 0.25 for some of the age groups. In these instances we excluded the risk factor from the model for that group. The C statistics ranged from 0.63 to

0.79, indicating discrimination typical for these types of models when predicting nursing homes' outcomes.^{13, 35-37}

Table 5 reports the results for a comparison of the QMs and EQMs for the 4th quarter of 2005 for all facilities that met the minimum sample size criterion (i.e. a denominator of at least 30 for the long-stay measures and at least 20 for the short-stay measure). Comparisons for other quarters, as well as comparisons within each state yielded very similar results and are not reported here.

The table presents information regarding the agreement about outlier designation between the QMs and the EQMs, for outliers defined as those at the 5, 10 and 20 percent tails of the distribution. Kappa values ranged from 0.80 to 0.90 for high quality outliers and from 0.63 to 0.86 for low quality outliers, indicating a moderate level of agreement. Kappa values tended to increase, indicating stronger agreement, as the definition of outliers was expanded to include a larger proportion of the distribution, going from the 5th percentile to the 20th. The error rates, both positive and negative, showed a similar trend. They ranged from a low of 7.9% to a high of 36.9%, with lower error rates for high quality rather than low quality outliers and for outliers defined as the 20th percentile compared with the 5th percentile. The higher agreement between the CMS QMs and the EQMs on high quality outliers than on low quality outliers may be an artifact of the censoring of outcome rates at zero, which affects the high quality tail of the distribution but not the low quality tail. Recall that all five measures are defined as rates of adverse outcomes, such that lower rates indicate higher quality.

Discussion

CMS is currently publishing 19 QMs that cover many aspects of the clinical care that nursing home residents receive. This information is expected to inform quality improvement efforts in nursing homes, to influence consumer demand, and to create market incentives for high quality care.^{1, 3, 38} In the future such measures might be used in pay for performance (P4P) programs.^{39, 40}

The study presented here assesses the impact of more extensive risk adjustment on the QMs published by CMS. It finds that there are additional risk factors beyond those included in the CMS methodology that are significantly associated with each of the five health outcomes studied. More importantly, it also finds that relative quality rankings change when these additional risk factors are added to the measures. While the Kappa statistics indicate moderate agreement on outlier classification, the error rates (when we view the EQMs as a gold standard – see further discussion below) range from about 10% to over 30%. Is this error rate acceptable for report cards that are expected to influence the behavior of consumers and providers? Is it acceptable for a P4P program?

Recent studies suggest that the Nursing Home Compare report card indeed has an important role to play. Pesis-Katz et al.⁴¹ find that consumers base their choice of a nursing home on the quality of the facility relative to other facilities in their local market. The study further demonstrates the important role that measures of clinical quality in particular might play in influencing the demand for nursing homes. Studies have also shown that a large percent of nursing homes take actions in response to their published scores in the Nursing Home Compare report⁴² and that these actions are associated with improvement in some of the measures.⁴³ This body of evidence, as well as the evidence on the impact of report cards in other areas of health care,^{44, 45} suggest that the accuracy of the information included in the Nursing Home Compare report cards is potentially very important.

The accuracy of the report cards will gain further importance if CMS, or other payers, implement a P4P program for nursing homes. Encouraged by the results of a pilot study

demonstrating that P4P improved the quality of hospital care CMS is planning to launch a three-year P4P nursing home demonstration in fifty facilities located in 4-5 states. While the design of the CMS P4P for nursing homes is not known at this time, it is likely to follow other programs, such as the Medicare P4P demonstration for hospitals which rewards only those in the top 10th and 20th percentile. Thus the accuracy of assignment of nursing homes to quality deciles would be important.

In addition to concerns about the fairness to providers and provision of “correct” market incentives, insufficient risk adjustment may actually create perverse incentives to avoid admission of the sickest patients, limiting access to nursing home care for those who need it the most. A national survey of nursing home administrators found that they are aware of issues related to inadequate risk adjustments, with 81% stating that the Nursing Home Compare QMs may be sensitive to unaccounted differences in case mix.⁴² Preliminary findings suggest that at least some nursing homes may be “cream skimming”, having changed the type of patients they admit following the publication of the Nursing Home Compare.⁴⁶ Evidence from a study of a more established report card, the NYS Cardiac Surgery Report, suggests that interventional cardiologists in NYS indeed are avoiding the sickest patients because of their concern about their ranking in the report card.⁴⁷

Concerns about the readiness of the QMs for prime time, either in a report card or as part of a P4P program, have been expressed by both the General Accounting Office⁴⁸ and the Institute of Medicine.⁴⁹ The analyses we present here lend support to these concerns.

Several other issues need to be considered as well. Our analyses focused on the adequacy of the risk adjustment. However, other aspects relevant to the construction of quality measures, including appropriate sample size and measurement errors, time window for measurement, and specification of the risk adjustment models, may be important. Furthermore, our departure points in developing the EQMs were the outcomes and denominators as defined by the QMs. The advantage of this is that it allows us to assess the incremental impact of risk adjustment. However, the EQMs might be improved further by adjusting the definition of the denominators as well.

A related issue is the validation of the EQMs. In calculating the error rate we assumed that the EQMs are the gold standard, and thus measured the QMs against the EQMs. However, we have not validated the EQMs across all possible dimensions. The EQMs have better face and content validity than the QMs.⁷ Their predictive validity (i.e. the C static) is similar to that found in other studies of nursing home outcomes. But we were unable to assess attributional and construct validity. Thus, while the EQMs cannot be considered a gold standard, they are an appropriate yardstick against which to assess the QMs to determine if they can be improved.

In summary, we have demonstrated in this study that quality ranking of nursing homes changes when risk adjustment is added to the measures used by CMS in its report card. Because these report cards have an impact on consumer choice and nursing homes performance the stakes are high and the appropriateness of the QMs should be investigated further.

Acknowledgments

The authors gratefully acknowledge funding from the National Institutes of Aging, Grant# AG023177 and #AG029608.

References

1. Mor V. Improving the quality of long-term care with better information. *Milbank Quarterly* 2005;83 (3):333–364. [PubMed: 16201996]

2. Winzelberg GS. The quest for nursing home quality: learning history's lessons. *Archives of Internal Medicine* 2003;163(21):2552–2556. [PubMed: 14638554]
3. Mukamel DB, Spector WD. Quality Report Cards and Nursing Home Quality. *The Gerontologist* 2003;43(special issue 2):58–66. [PubMed: 12711725]
4. Centers for Medicare & Medicaid Services. Nursing Home Quality Initiatives. [Dec 1, 2006]. webpage]. 12/1/06; General Information. Available at: <http://www.cms.hhs.gov/QualityInitiativesGenInfo/>
5. Harrington C, Collier E, O'Meara J, Kitchener M, Simon LP, Schnelle JF. Federal and state nursing facility websites: just what the consumer needs? *Am J Med Qual* 2003;18(1):21–37. [PubMed: 12583642]
6. Berlowitz DR, Rosen A, Wang F, et al. Purchasing or Providing Nursing Home Care: Can Quality Data Provide Guidance. *Journal of the American Geriatric Society* Apr;2005 53(4):603–608.
7. Iezzoni, LI. Risk Adjustment for Measuring Health Care Outcomes. Ann Arbor, Michigan: Health Administration Press; 1994.
8. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjustment Method. *American Journal of Public Health* 1996;86:1379. [PubMed: 8876505]
9. Glance LG, Osler TM, Dick AW. Evaluating trauma center quality: does the choice of the severity-adjustment model make a difference? *J Trauma* Jun;2005 58(6):1265–1271. [PubMed: 15995480]
10. Glance LG, Osler TM, Dick A. Rating the quality of intensive care units: is it a function of the intensive care unit scoring system? *Crit Care Med* Sep;2002 30(9):1976–1982. [PubMed: 12352029]
11. Mukamel DB, Brower CA. The Influence of Risk Adjustment Methods on Conclusions About Quality of Care in Nursing Homes Based on Outcome Measures. *The Gerontologist* 1998;38(6):695–703. [PubMed: 9868849]
12. Berlowitz DR, Ash AS, Brandeis GH, Brand HK, Halpern JL, Moskowitz MA. Rating long-term care facilities on pressure ulcer development: importance of case-mix adjustment. *Annals of Internal Medicine* Mar 15;1996 124(6):557–563. [PubMed: 8597318]
13. Rosen A, Wu J, Chang BH, et al. Risk adjustment for measuring health outcomes: an application in VA long-term care. *Am J Med Qual* Jul-Aug;2001 16(4):118–127. [PubMed: 11477956]
14. Mor V. Defining and measuring quality outcomes in long-term care. *J Am Med Dir Assoc* Mar;2007 8(3 Suppl 2):e129–137. [PubMed: 17352977]
15. Mukamel DB, Dick A, Spector WD. Specification Issues in Measurement of Quality of Medical Care Using Risk Adjusted Outcomes. *Journal of Economic and Social Measurement* 2000;26:267–281.
16. Li, Y.; Dick, A.; Glance, LG.; Cai, X.; Mukamel, D. Misspecification Issues in Risk Adjustment and Construction of Outcome-Based Quality Indicators. State University of New York; Buffalo: 2006. p. 28
17. Arling G, Lewis T, Kane RL, Mueller C, Flood S. Improving Quality Assessment through Multilevel Modeling: The Case of Nursing Home Compare. *Health Serv Res* Jun;2007 42(3p1):1177–1199. [PubMed: 17489909]
18. Jha AK, Epstein AM. The predictive accuracy of the New York State coronary artery bypass surgery report-card system. *Health Affairs (Millwood)* May-Jun;2006 25(3):844–855.
19. Zimmerman D. Resident based outcome measures: their potential in nursing home quality improvement [abstract]. *AHSR & FHSR Annual Meeting Abstract Book* 11:57X 1997:57X.
20. Zimmerman DR, Karon SL, Arling G, et al. Development and Testing of Nursing Home Quality Indicators. *Health Care Financing Review* 1995;16:107. [PubMed: 10151883]
21. Morris, JN.; Moore, T.; Jones, R., et al. Validation of Long-Term and Post-Acute Care Quality Indicators. Cambridge MA: Abt Associates Inc.; Jun 10. 2003 CMS Contract No: 500-95-0062 / T.O.#4
22. Berg, K.; Fries, BE.; Jones, R., et al. Identification and Evaluation of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings. Abt Associates Inc.; Oct 1. 2001 CMS Contract No: 500-95-0062 / T.O. #4
23. Centers for Medicare & Medicaid Services. Nursing Home Quality Initiative Overview. [8/30/07]. (Updated 1/20/04). <http://www.cms.hhs.gov/NursingHomeQualityInits/downloads/NHQIOverview.pdf>

24. National Quality Forum. National voluntary consensus standards for nursing home care: A consensus report. Vol. NQFCR-06-04. Washington DC: 2004.
25. Harris Y, Clauser SB. Achieving improvement through nursing home quality measurement. *Health Care Financ Rev Summer*;2002 23(4):5–18. [PubMed: 12500467]
26. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the Minimum Data Set for nursing home resident assessment and care screening (MDS). *Gerontologist* 1995;35:172. [PubMed: 7750773]
27. Mor V, Angelelli J, Jones R, Roy J, Moore T, Morris J. Inter-rater reliability of nursing home quality indicators in the U.S. *BMC Health Serv Res* 2003;3(1):20. [PubMed: 14596684]
28. Morris JN, Hawes C, Fries BE, et al. Designing the National Resident Assessment Instrument for Nursing Homes. *The Gerontologist* 1990;30:293. [PubMed: 2354790]
29. Abt Associates Inc.. National Nursing Home Quality Measures: User's Manual, November, 2004 (v1.2). Cambridge, MA: U.S. Department of Health & Human Services, Center for Medicare & Medicaid Services; 2004.
30. Hosmer, DW.; Lemeshow, S. Applied Logistic Regression. New York NY: John Wiley & Sons; 1989.
31. Mukamel DB, Watson NM, Meng H, Spector WD. Development of a Risk-Adjusted Urinary Incontinence Outcome Measure of Quality for Nursing Homes. *Medical Care* 2003;41(4):467–478. [PubMed: 12665711]
32. Li Y, Dick A, Glance LG, Cai X, Mukamel D. Misspecification Issues in Risk Adjustment and Construction of Outcome Based Quality Indicators. *HSOR Jun*;2007 7(12):39–56.
33. Altman, DG. Practical Statistics for Medical Research. Vol. 7. Chapman & Hall. CRC; 1991.
34. Glance LG, Dick AW, Osler TM, Mukamel DB. Accuracy of hospital report cards based on administrative data. *Health Serv Res Aug*;2006 41(4 Pt 1):1413–1437. [PubMed: 16899015]
35. Mukamel DB. Risk Adjusted Outcome Measures and Quality of Care in Nursing Homes. *Medical Care April*;1997 35(4):367–385. [PubMed: 9107205]
36. Berlowitz DR, Brandeis GH, Anderson JJ, et al. Evaluation of a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc Jul*;2001 49(7):872–876. [PubMed: 11527477]
37. Berlowitz DR, Brandeis GH, Morris JN, et al. Deriving a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc Jul*;2001 49(7):866–871. [PubMed: 11527476]
38. Harrington C, O'Meara J, Kitchener M, Simon LP, Schnelle JF. Designing a report card for nursing facilities: what information is needed and why. *Gerontologist* 2003;43(Spec No 2):47–57. [PubMed: 12711724]
39. Centers for Medicare & Medicaid Services. Medicare “Pay for Performance (P4P)” Initiatives. [June 20, 2007]. <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1343>
40. Abt Associates Inc. Nursing Home Pay-for-Performance Demonstration Draft Design. [June 20, 2007]. http://www3.cms.hhs.gov/DemoProjectsEvalRpts/downloads/NHP4P_Handout.pdf
41. Pesis-Katz, I.; Phelps, CE.; Temkin-Greener, H.; Spector, WD.; Mukamel, D. AcademyHealth. Orlando, FL: 2007. Why are Nursing Homes' Quality Report Cards Important? Evidence from California, Ohio, New York and Texas.
42. Mukamel D, Spector WD, Zinn JS, Huang L, Weimer D, Dozier A. Nursing Homes' Response to the Nursing Home Compare Report Card. *J Gerontol B Psychol Sci Soc Sci Jul*;2007 62(4):S218–225. [PubMed: 17673535]
43. Mukamel, D.; Weimer, D.; Ladd, H.; Spector, WD.; Zinn, JS. AcademyHealth. Orlando, FL: 2007. Actions taken by nursing homes in response to publication of the Nursing Home Compare report card.
44. Mukamel DB, Weimer DL, Zwanziger J, Huang-Gorthy S, Mushlin AI. Quality report cards, selection of cardiac surgeons and racial disparities: A study of the publication of the NYS Cardiac Surgery Reports. *Inquiry* 2004;2005;41(4):435–446. [PubMed: 15835601]
45. Mukamel D, Weimer D, Mushlin A. Interpreting Market Share Changes as Evidence for Effectiveness of Quality Report Cards. *Med Care*. 2007 in press

46. Mukamel, D.; Spector, WD.; Weimer, D.; Zinn, JS.; Li, Y. AcademyHealth. Orlando, FL: 2007. Did the publication of the Nursing Home Compare report card lead to “cream skimming” by nursing homes?.
47. Narins CR, Dozier AM, Ling FS, Zareba W. The influence of public reporting of outcome data on medical decision making by physicians. Arch Intern Med 2005;165(1):83. [PubMed: 15642879]
48. U. S. General Accounting Office. Nursing Homes: Public Reporting of Quality Indicators Has Merit, but National Implementation Is Premature GAO-03-187. Washington D.C.: U.S. General Accounting Office; 2002.
49. Institute of Medicine. Committee on Redesigning Health Performance. Washington D.C.: 2007.

Table 1
Definitions of the CMS Quality Measures (QMs)

Quality Measures	Exclusion Criteria [*]	Covariate Risk Adjusted
Percent of long-stay residents:		
Given Influenza Vaccination During the Flu Season	(9) (10)	No
Who Were Assessed and Given Pneumococcal Vaccination	(10)	No
Whose Need for Help With Daily Activities Has Increased	(2) (3) (4) (5)	No
Who Have Moderate to Severe Pain	(1) (2)	Independence or modified decision-making ability on prior assessment
Who have pressure ulcers (high-risk) ²³	(1) (2) (6)	No
Who have pressure ulcers (low-risk) ²³	(1) (2) (6)	No
Who were physically restrained	(1) (2)	No
Who are more depressed or anxious	(4) (11) (12)	No
Who lost control of their bowels or bladder (low-risk)	(1) (2) (4) (13) (14) (15)	No
Who have/had a catheter inserted and left in their bladder	(1) (2)	Bowel incontinence, pressure ulcers on the prior assessment
Who spent most of their time in bed or in a chair	(1) (2) (4)	No
Whose ability to move about in and around their room got worse	(4) (5) (16)	Recent falls, extensive support or more dependence in eating / toileting on the prior assessment
With an urinary tract infection	(1) (17)	No
Who lost too much weight	(1) (18)	No
Percent of short-stay residents:		
Given influenza vaccination during the flu season	(9) (10)	No
Who were assessed and given pneumococcal vaccination	(10)	No
With delirium	(2) (4) (5)	Prior residential history
Who had moderate to severe pain	(7)	No
With pressure ulcers	(8)	History of resolved pressure ulcers, limited assistance in bed mobility, bowel incontinence, diabetes or peripheral vascular disease, and Low Body Mass Index

* Exclusion criteria:

- (1) target assessment is an admission;
- (2) the QM did not trigger (resident not in numerator) and/or there is missing or inconsistent data on Minimum Data Set (MDS) items required for quality measure (QM);
- (3) MDS item cannot show decline because it has total dependence value if activity did not occur;
- (4) resident is comatose or comatose status is unknown;
- (5) resident has end-stage disease or receives hospice care;
- (6) not qualify as high-risk and bed mobility or transfer or comatose is unknown;
- (7) pain symptoms are unknown or inconsistent on 14-day assessment;
- (8) pressure ulcers are missing in the 5-day / 14-day assessment;
- (9) not in facility during influenza season or facility unable to obtain vaccine;
- (10) not eligible for given vaccine or resident offered vaccine but declined;
- (11) the Mood Scale score is missing on the target or prior assessment;
- (12) the Mood Scale score is at a maximum (value 8) on the prior assessment;
- (13) high risk residents with severe cognitive impairment or totally dependent in mobility ADLs;
- (14) resident has an indwelling catheter or an ostomy;
- (15) not qualify as high risk and the cognitive impairment items or any of the mobility ADLs unknown;
- (16) locomotion on unit is unknown or shows some dependence on target or prior assessment;

- (17) urinary tract infection in last 30 days is missing on target assessment;
- (18) weight loss is missing or resident is receiving hospice care or hospice status unknown;

Table 2
Distribution of QMs (Data are shown for 4th quarter of 2005)

QM	Number of Nursing Homes with Non Missing QMs	Average	Standard Deviation	Coefficient of Variation [*] (%)	Range		
					5 th Percentile	Median	95 th Percentile
% of Long-Stay residents with:							
Pressure Ulcers-High risk patients	9,971	12.8	7.2	56	2.9	11.9	25.7
Pressure Ulcers-Low risk patients	6,732	2.4	2.8	117	0.0	2.0	7.7
Physical Restraints	13,639	6.6	7.6	116	0.0	4.2	21.2
Pain	13,408	8.4	7.5	90	0.0	6.5	22.7
% of Short-Stay residents with :							
Pain	9,792	23.6	14.1	60	4.5	21.4	50.0

^{*} Defined as the ratio of the standard deviation to the average times 100.

Table 3
Descriptive Statistics (Standard deviations in parentheses)*

		Mean
QMs/outcome	% of residents with pressure ulcers in high risk (long-stay)	13.7
	% of residents with pressure ulcers in low risk (long-stay)	2.6
	% of residents in physical restraints (long-stay)	8.2
	% of residents with pain (long-stay)	9.4
	% of residents with pain (short-stay)	24.7
Risk factors (long-stay patients)	Age	80.6(13.3)
	Female (%)	71.8
	Making self understood	
	Sometimes (%)	16.9
	Rarely (%)	11.1
	Ability to understand others	
	Sometimes (%)	20.6
	Rarely (%)	8.7
	Behavior problems (<u>severity scale ranges from 0-3</u>)	
	Wandering	0.11(0.51)
	Verbally abusive	0.09(0.40)
	Physically abusive	0.06(0.32)
	Socially inappropriate	0.16(0.57)
	Resists care	0.25(0.68)
	ADL self-performance	
	Limited assistance in bed mobility (%)	15.8
	Dependence in bed mobility (%)	45.0
	Limited assistance in transfer (%)	16.9
	Dependence in transfer (%)	53.7
	Limitation in bed mobility and transfer (%)	59.4
	Number of pressure ulcers <u>per resident</u>	
	Stage 1	0.04(0.26)
	Stage 2	0.10(0.46)
	Stage 3	0.02(0.20)
	Stage 4	0.03(0.27)
	Fall in last 180 days (%)	34.8
	Depression (%)	14.8
	Loss of voluntary movement in arm, hand, foot or leg (%)	11.7
	Bedfast (%)	4.7
	Incontinence (bowel or bladder) (%)	59.6
	Weight loss of at least 5% in 30 days or at least 10% in 180 days (%)	9.6
	Cognitive skills in decision making on prior assessment (%)	38.0
Risk factors (short-stay patients)	Age	80.4(10.7)
	Female (%)	66.7

	Mean
Making self understood	
Sometimes (%)	9.4
Rarely (%)	4.0
Ability to understand others	
Sometimes (%)	11.7
Rarely (%)	3.0
Number of pressure ulcers <u>per resident</u>	
Stage 1	0.15(0.54)
Stage 2	0.27(0.78)
Stage 3	0.05(0.33)
Stage 4	0.09(0.48)
Specific diseases	
Hip fracture (%)	11.2
Cerebral palsy (%)	0.2
Hemiplegia/hemiparesis (%)	8.0
Wounds infection (%)	0.1
Skin condition	
Skin abrasions or bruises (%)	42.7
Burns (%)	0.3
Skin tears or cuts (%)	8.4
Surgical wounds (%)	25.8
Infection of the foot (%)	1.8
Fall in last 180 days (%)	47.1
Depression (%)	14.5

* Standard deviation is not reported for dichotomous variables

Note: This table presents all the risk factors considered in the initial models. Only those that were significant at the 0.25 level or better were included in the final models, shown in table 4. Exclusion of risk factors was confirmed by performing an F test of the hypothesis that all the excluded variables were jointly insignificantly different from zero.

Table 4
Risk Adjustment Models Predicting the Log Odds of Outcomes by Age Group: Table reports odds ratios.

Model/Outcome	Risk factors	Age <=64	Age 65-74	Age 75-84	Age >=85
Model 1: Pressure Ulcers (high risk) (long-stay patients)	Age	1.01**			
	Female	0.84**	0.84**	0.75**	0.67**
	ADL in bed mobility and transfer:				
	Limited assistance in bed mobility	1.43*	1.45	1.46**	1.48*
	Dependence in bed mobility	1.95**	2.37**	2.43**	2.47**
	Dependence in transfer			1.89**	1.81**
	Loss of voluntary movement	1.78**	1.36**	1.36**	1.44**
	Bedfast	2.51**	2.69**	2.78**	2.73**
	Weight loss	2.43**	2.54**	2.55**	2.47**
	Sample size	6,708	6,869	19,965	30,985
	C statistics	0.68	0.66	0.66	0.65
Model 2: Pressure Ulcers (low risk) (long-term patients)	Age	0.98*			
	Female			0.72**	0.64**
	ADL in bed mobility and transfer:				
	Limited assistance in bed mobility		1.51*	1.55**	1.57**
	Limited assistance in transfer	2.26**	2.41**	2.36**	2.12**
	Loss of voluntary movement	5.39**	1.72	1.51	
	Bedfast	3.85*		2.36	2.05
	Weight loss	2.29*	2.69**	2.62**	2.57**
	Sample size	6,659	6,469	15,706	22,503
	C statistics	0.73	0.71	0.68	0.68
Model 3: Physical Restraints (long-stay patients)	Age	0.98**			0.99
	Making self understood: Always or usually				Reference category

Model/Outcome	Risk factors	Age <=64	Age 65-74	Age 75-84	Age >=85
	Sometimes	2.97**	2.09**	1.83**	1.79**
	Rarely	4.47**	2.73**	2.22**	2.02**
	Behavior problems:				
	Physically abusive	1.25*	1.24*	1.22**	1.20**
	Socially inappropriate	1.12*	1.14*	1.16**	1.21**
	ADL limitation in bed mobility	1.74**	1.94**	1.95**	1.80**
	Fall in last 180 days	1.24*	1.49**	1.48**	1.36**
	Ability to understand others:				
	Always or usually			Reference category	
	Sometimes	1.57**	2.33**	2.56**	2.41**
	Rarely	1.96**	3.03**	3.09**	2.85**
	Sample size	12,998	13,185	35,810	53,541
	C statistics	0.78	0.77	0.74	0.72
Model 4: Pain (long-stay patients)	Age			0.98*	0.98**
	Gender	1.44**	1.69**	1.60**	1.42**
	Making self understood:				
	Always or usually			Reference category	
	Sometimes	0.68*	0.72*	0.76**	0.79**
	Rarely	0.61*	0.55**	0.60*	0.60**
	Ability to understand others:				
	Always or usually			Reference category	
	Sometimes	0.75		0.77**	
	Rarely	0.58		0.67*	
	Sometimes or rarely		0.73*		0.88
	Depression	1.67**	1.72**	1.75**	1.74**
	Fall in last 180 days	1.19**	1.17**	1.18**	1.22**
	Number of pressure ulcers in:				

Model/Outcome	Risk factors	Age <=64	Age 65-74	Age 75-84	Age >=85
	Stage 1	1.29*	1.33*	1.29**	1.27**
	Stage 2	1.36**	1.33**	1.33**	1.34**
	Stage 3	1.52**	1.55*	1.59**	1.67**
	Stage 4	1.59**	1.55**	1.56**	1.61**
	Cognitive skills in decision making on previous assessment	2.61**	2.16**	2.10**	1.98**
	Sample size	12,790	13,317	35,575	53,562
	C statistics	0.68	0.67	0.66	0.63
Model 5: Pain (short-stay patients)	Age		0.97**	0.96**	0.98**
	Female	1.50**	1.74**	1.66**	1.51**
	Making self understood:				
	Always or usually			Reference category	
	Sometimes	0.42**	0.54**	0.61**	0.72*
	Rarely	0.28*	0.36*	0.45**	0.55**
	Ability to understand others:				
	Always or usually			Reference category	
	Sometimes		0.54**	0.59**	
	Rarely		0.52	0.56*	
	Sometimes or rarely	0.49**			0.70**
	Depression	1.30**	1.28**	1.29**	1.37**
	Fall in last 180 days			1.16**	1.34**
	Number of pressure ulcers in:				
	Stage 2	1.10*	1.08*	1.08**	1.08**
	Stage 3	1.16	1.15*	1.18**	1.20**
	Stage 4	1.14*	1.14*	1.17**	1.21**
	Hip fracture	1.55*	1.48**	1.41**	1.52**
	Skin condition:				
	Skin abrasions or bruises	1.28**	1.30**	1.25**	1.26**

Model/Outcome	Risk factors	Age <=64	Age 65-74	Age 75-84	Age >=85
	Surgical wounds	2.50**	2.54**	2.40**	2.15**
	Infection of the foot			1.30*	1.48**
	Sample size	3,672	8,334	20,764	20,659
	C statistics	0.69	0.69	0.69	0.68

* 0.01 <=p < 0.05;
** p < 0.01

Table 5
Agreement between CMS QMs and Risk-Adjusted EQMs on identification of outliers

Part I. Outliers defined as the top and bottom 5% of the distribution						
High-Quality Outliers			Low Quality Outliers			
Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	
Long-Stay QMs:						
Pressure Ulcers-High risk patients	0.88	13.5				
Pressure Ulcers-Low risk patients	n/a ^b	n/a ^b	0.66	32.3	32.3	
Physical Restraints	n/a ^b	n/a ^b	0.76	22.6	23.7	
Pain	n/a ^b	n/a ^b	0.71	27.9	27.9	
Short-stay QMs:						
Pain	0.80	19.4	0.75	23.6	23.6	
			0.63	33.5	36.9	
Part II. Outliers defined as the top and bottom 10% of the distribution						
High-Quality Outliers			Low Quality Outliers			
Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	
Long-Stay QMs:						
Pressure Ulcers-High risk patients	0.90	7.9				
Pressure Ulcers-Low risk patients	n/a ^b	n/a ^b	0.72	24.7	25.1	
Physical Restraints	n/a ^b	n/a ^b	0.78	17.7	22.3	
Pain	0.86	12.6	0.75	22.0	22.5	
Short-stay QMs:						
Pain	0.84	14.4	0.79	18.6	18.6	
			0.70	26.1	28.0	
Part III. Outliers defined as the top and bottom 20% of the distribution						
High-Quality Outliers			Low Quality Outliers			
Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	
Long-Stay QMs:						
Pressure Ulcers-High risk patients	0.88	9.3				
Pressure Ulcers-Low risk patients	n/a ^b	n/a ^b	0.78	17.8	17.9	
Physical Restraints	n/a ^b	n/a ^b	0.86	11.3	11.7	
			0.82	14.4	14.8	

Part I. Outliers defined as the top and bottom 5% of the distribution

	High-Quality Outliers			Low Quality Outliers		
	Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)	Kappa	False negative error rate ^a (%)	False positive error rate ^a (%)
Pain	0.84	12.0	13.4	0.83	13.9	14.0
Short-stay QMs:						
Pain	0.86	11.1	11.2	0.77	18.5	18.7

^aFalse negative error rate = 100 (1- sensitivity). False positive error rate = 100 (1- positive predictive values).

^bSince more than 5% (10% or 20%) of the nursing homes had a prevalence of zero, high quality outliers could not be determined based on our definition.