

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2009.01061.x
RESEARCH ARTICLE

The “Nursing Home Compare” Measure of Urinary/Fecal Incontinence: Cross-Sectional Variation, Stability over Time, and the Impact of Case Mix

Yue Li, John Schnelle, William D. Spector, Laurent G. Glance, and Dana B. Mukamel

Objectives. To assess the impact of facility case mix on cross-sectional variations and short-term stability of the “Nursing Home Compare” incontinence quality measure (QM) and to determine whether multivariate risk adjustment can minimize such impacts.

Study Design. Retrospective analyses of the 2005 national minimum data set (MDS) that included approximately 600,000 long-term care residents in over 10,000 facilities in each quarterly sample. Mixed logistic regression was used to construct the risk-adjusted QM (nonshrinkage estimator). Facility-level ordinary least-squares models and adjusted R^2 were used to estimate the impact of case mix on cross-sectional and short-term longitudinal variations of currently published and risk-adjusted QMs.

Principal Findings. At least 50 percent of the cross-sectional variation and 25 percent of the short-term longitudinal variation of the published QM are explained by facility case mix. In contrast, the cross-sectional and short-term longitudinal variations of the risk-adjusted QM are much less susceptible to case-mix variations (adjusted $R^2 < 0.10$), even for facilities with more extreme or more unstable outcome.

Conclusions. Current “Nursing Home Compare” incontinence QM reflects considerable case-mix variations across facilities and over time, and therefore it may be biased. This issue can be largely addressed by multivariate risk adjustment using risk factors available in the MDS.

Key Words. Quality measure, Nursing Home Compare, incontinence, case mix, stability

Ensuring and improving the quality of nursing home (NH) care is a primary concern of the federal and state governments, policy makers, and consumers (Institute of Medicine 2001). In addition to government regulations which set minimum criteria of care (Harrington, Mullan, and Carrillo 2004), and the

national quality improvement activities by Quality Improvement Organizations (General Accounting Office [GAO] 2007), NH quality report cards have played an important role in the national strategy of market-driven quality improvement (GAO 2002; Mukamel et al. 2008a; Li et al. 2009). Currently, the Centers for Medicare and Medicaid Services (CMS) maintain a website entitled “Nursing Home Compare” that publishes outcome measures derived from resident health assessments. These measures are designed to provide quality benchmarks to potential NH consumers and to inform their choice of facilities (GAO 2002). These quality measures (QMs) can also serve as objective evidence of performance and enable pay-for-performance (P4P) programs to reward facilities with superior outcomes (Abt 2006). Recent studies suggest that NHs may respond to the CMS reports and take actions to improve practices affecting the published QMs (Mukamel et al. 2007; Mukamel et al. 2008b).

To maximize the value of NH QMs, it is important that they reflect true performance differences between facilities. However, facility variation in QM rates may comprise both varying case mix (i.e., variation in the patient populations) and varying care practices (i.e., variation in performance). It is thus necessary that NH QMs be risk adjusted for resident frailties and functional impairment so that the impact of case mix on QM rates is minimized (Iezzoni 2003).

The CMS QMs are only minimally risk adjusted, mostly through exclusion criteria (Mukamel et al. 2008a). Recent research reports that these QMs show quite different results in outcome rankings than the measures that incorporate further statistical risk adjustment (Arling et al. 2007; Mukamel et al. 2008a; Li et al. 2009). In addition, a study (Simmons et al. 2003) found that a selected group of Southern California NHs with extremely high or low rates in the weight-loss QM did not differ in their nutrition care practices; rather, the difference in QM rate across facilities was dominated by differences in case mix.

Address correspondence to Yue Li, Ph.D., Division of General Internal Medicine, SE610 GH, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242; e-mail: yue-li@uiowa.edu. John Schnelle, Ph.D., is with the Vanderbilt Center for Quality Aging, Vanderbilt University, Nashville, TN. William D. Spector, Ph.D., is with the Center for Delivery, Organization, and Markets, Agency for Healthcare Research and Quality, Rockville, MD. Laurent G. Glance, M.D., is with the Department of Anesthesiology, the University of Rochester School of Medicine and Dentistry, Rochester, NY. Dana B. Mukamel, Ph.D., is with the Department of Medicine, Health Policy Research Institute, University of California, Irvine, CA.

This body of literature has focused on the cross-sectional variation of QMs and the impact of case mix on it. Another desired, although less tested, property of the QMs is that they should be relatively stable when calculated over a short period of time (Berlowitz et al. 1998; Karon, Sainfort, and Zimmerman 1999; Rantz et al. 2004). Karon, Sainfort, and Zimmerman (1999) report that with several exceptions, a facility’s outcome rates tend to show high stability (or correlation) over a 3-month or 6-month period. In general, although QM rates may change in the longer term as a result of changing facility practices, staff turnover, or altered market conditions, the short-term stability of QMs is expected and bears important implications for the “Nursing Home Compare” so that it can provide relevant and reliable performance information to support consumer and policy decisions. It is unknown, however, whether facility case mix empirically affects the short-term stability of facility’s QM rate.

STUDY HYPOTHESES AND SIGNIFICANCE

This study expanded prior research and had two objectives: (1) to assess the potential impact of facility case mix on both the cross-sectional variations and short-term stability of the CMS QM and (2) to determine whether multivariate risk adjustment for resident characteristics can minimize these impacts. Theoretically, individual health outcomes are determined by both intrinsic patient risks and the quality of care the patient receives (Iezzoni 2003). Because aggregate-level case mix may not be evenly distributed across facilities or longitudinally within a facility, a facility’s outcome rate, when unadjusted or minimally risk-adjusted, would be affected by case mix in both ways. Therefore, the first hypothesis we tested is as follows:

Hypothesis 1: Facility case mix explains a substantial portion of both the cross-sectional variation and the short-term longitudinal variation of the CMS QM.

In addition, previous studies (Mukamel et al. 2008a; Li et al. 2009) suggested that NHs with extremely high or low QM rates are more likely than others to be affected by case-mix adjustment. For example, the cross-sectional rankings of these “extreme” facilities showed a higher level of discrepancy when rankings were derived from the CMS QM versus further risk-adjusted QM, suggesting a potentially higher influence of case mix on these extreme facilities. Therefore, we further hypothesized as follows:

Hypothesis 2: The impact of case mix is even larger for NHs with extreme QM rate or with extreme change of QM rate over the short term.

Finally, multivariate risk adjustment is an approach to balancing the effect of patient risks on outcomes, before measuring “quality” based on residual outcome differences across facilities (Iezzoni 2003). We thus expected that multivariate risk adjustment could remove or at least partially reduce the role of facility case mix in explaining the cross-sectional and longitudinal variation of facility outcome. This motivated our final hypothesis as follows:

Hypothesis 3: Multivariate risk adjustment of the QM can reduce the impact of facility case mix on both the QM’s cross sectional variation and its short-term variability, even for facilities with more extreme performance.

We note that current risk-adjustment models are typically performed on patient-level cross-sectional data (Iezzoni 2003). Although studies in the acute care area suggested that such risk adjustment, when appropriately performed, can largely reduce the impact of case mix on the cross-sectional (or between-facility) variation in outcomes (Green, Passman, and Wintfeld 1991), no research is available to inform how risk adjustment can dampen the potential time-variant effect of case mix, that is, the effect on the QM’s short-term stability. The only relevant study for NH services is by Berlowitz et al. (1998) who reported similar levels of short-term stability between a facility’s unadjusted and risk-adjusted pressure ulcer rate. Nevertheless, that study did not examine the role of case mix in the stability, or within-facility variation, of the two types of outcome over time.

In this paper, we focused on one of the 19 QMs (Mukamel et al. 2008a) currently published by CMS—urinary/fecal incontinence for chronic care residents—and tested the impact of case mix on its cross-sectional variation and short-term stability. Incontinence, among other published outcomes, significantly affects the social and psychological well-being of NH residents. Meanwhile, incontinence is amenable to appropriate care such as effective behavioral or pharmacological treatments (AHRQ 1996) and can be improved by appropriate quality management approaches in NHs (Schnelle et al. 1993, 1995, 2002; Levy-Storms, Schnelle, and Simmons 2007). Therefore, NH incontinence rate can serve as an evidence-based indicator of long-term care performance.

METHODS

Data Source

This study involved retrospective analyses of the 2005 national minimum data set (MDS) that includes all Medicare and/or Medicaid certified NHs (over 90 percent of all NHs nationally) whose QMs are being published by CMS. The MDS has over 350 data elements containing demographic, functional, and clinical information on individual residents. MDS assessments for long-term care residents are performed by facility staff upon admission, quarterly thereafter, and when the resident has a significant change of health status. Validation studies (Hawes et al. 1995; Lawton et al. 1998; Mor et al. 2003) have shown that MDS records, especially those used for calculating the QMs, meet general criteria for reliability and accuracy.

The CMS QM

We focused on one published QM for chronic care residents—percent of the low-risk residents who lose control of bowels or bladder. According to CMS’s definition, “low-risk” residents exclude those with severe cognitive impairment, with total dependence in mobility activities of daily living (ADLs), in coma, or with an indwelling catheter (Abt 2004). CMS also excludes facilities that have fewer than 30 eligible residents in the published QM. We followed these exclusion criteria to first define a binary variable y_{ij} for eligible resident i in facility j that equals 1 if the resident showed frequent or full incontinence, and zero otherwise. We then calculated the CMS QM (O_j) for each NH as the number of low-risk residents with incontinence (i.e., $y_{ij} = 1$) divided by n_j , the total number of low-risk residents in the facility. For each facility we calculated a set of three QMs using MDS of the first, second, and third quarter of 2005 separately.

The Risk-Adjusted QM

Selection of Risk Factors. We examined all variables in the MDS and identified potential predictors of incontinence in NH residents based on previous literature (Ouslander and Schnelle 1995; Nelson, Furner, and Jesudason 1998; Chassagne et al. 1999; Mukamel et al. 2003; Schnelle and Leung 2004; Nelson and Furner 2005) and clinical considerations (see the supporting information Appendix SA2 for the list of potential risk factors). Earlier studies have developed or reviewed risk-adjustment models for urinary incontinence (Mukamel et al. 2003; Nelson and Furner 2005), fecal incontinence (Chassagne et al. 1999; Nelson and Furner 2005), or both (Ouslander and

Schnelle 1995; Nelson and Furner 2005), and been instrumental to our identification of potential predictors. In bivariate analyses (χ^2 test or t -test) we confirmed that all candidate predictors showed meaningful associations with the dependent variable y_{ij} at the significance level of 0.001.

In multivariate analyses, each quarterly sample was randomly split into two halves (the development sample and the validation sample), before multivariate logistic regression models of y_{ij} were estimated on each split sample for the purpose of cross-validation. We tested main effects and possible interactions between potential risk factors and excluded multivariate predictors that were insignificant at the 0.001 level. When building the cross-validated model we also considered two factors. First, we avoided adjusting for the variables that predicted the outcome but reflected essentially NH practices, especially practices related to continence care, because including these “endogenous” predictors for risk adjustment would deflate facilities’ actual outcome variations that we intended to quantify. This issue of “over-adjustment” has been described in recent studies (Elliott et al. 2001; Mukamel et al. 2008a; Li et al. 2009), and we will return to this issue in the “Discussion.”

Second, we confirmed that the risk factors that entered the model varied across NHs. This verification is important in the context of risk adjustment because a characteristic that is unevenly distributed over facilities (e.g., facility gender mix) represents a potential source of bias in outcome comparisons and thus warrants statistical adjustment (O’Malley et al. 2005). On the other hand, a variable that predicts the outcome but is relatively homogeneous across facilities may not necessarily be an appropriate risk adjustor (O’Malley et al. 2005).

The final set of resident characteristics used for multivariate adjustment (see Table 1) included age (< 65 years, 65–74 years, 75–84 years, and \geq 85 years), female gender (yes/no), cognitive skill impairment (yes/no), delirium (yes/no), being able to make self understood (always or usually, sometimes, and rarely), presence of behavioral symptoms (wandering, and verbally abusive or socially inappropriate/disruptive behaviors), and limitations in ADLs (transfer, locomotion on unit, eating, toilet use, personal hygiene, and bathing). Each of the six ADL items ranged from 0 to 4 with 0 indicating total independence and 4 indicating total dependence.

Model Estimation and Construction of Risk-Adjusted QMs. The final risk-adjustment model was estimated on each entire quarterly data using SAS (SAS Corp., Cary, NC) Proc Glimmix that, to accommodate the clustering of

Table 1: Resident Characteristics ($n = 600,580$) and Estimates in the Risk-Adjustment Model: The Minimum Data Set of the First Quarter of 2005

Resident Characteristic	Prevalence (%) or Mean \pm SD	Risk Adjustment Model		
		β -Coefficient	Odds Ratio	p Value
Urinary and/or fecal incontinence	47.39			
Age (years)	80.78 \pm 13.07			
< 65	12.12	-0.510	0.600	.0000
65-74	11.81	-0.200	0.819	.0000
75-84	29.13	0.030	1.031	.0949
≥ 85	46.94		Reference	
Female	70.42	-0.051	0.950	.0003
Age-female interaction				
< 65 \times female		0.330	1.391	.0000
65-74 \times female		0.158	1.171	.0000
75-84 \times female		0.041	1.042	.0493
$\geq 85 \times$ female			Reference	
Cognitive skill impairment	53.66	0.621	1.861	.0000
Delirium	38.75	0.064	1.066	.0000
Make self-understood				
At least usually	87.00		Reference	
Sometimes	11.95	0.386	1.471	.0000
Rarely	1.05	0.699	2.012	.0000
Behavioral symptoms*	20.77	0.098	1.103	.0000
Physically abusive	4.41	0.273	1.312	.0000
Activities of daily living (ADLs)				
Transfer 0-4	1.83 \pm 1.40	0.078	1.081	.0000
Locomotion on unit 0-4	1.49 \pm 1.49	0.025	1.025	.0000
Eating 0-4	0.81 \pm 1.10	0.164	1.178	.0000
Toilet use 0-4	2.16 \pm 1.42	1.117	3.056	.0000
Personal hygiene 0-4	2.30 \pm 1.24	0.248	1.282	.0000
Bathing 0-4	2.92 \pm 0.95	0.265	1.304	.0000
Intercept		-4.854	—	.0000
c -statistic			0.914	

*Include wandering, and verbally abusive or socially inappropriate/disruptive behaviors.

residents in NHs, incorporates a random intercept u_j for each facility and fixed effects for resident predictors (Littell et al. 2006). The choice between alternative modeling approaches, such as mixed- versus fixed-effects models, was discussed before (Li et al. 2009). The “mixed” procedure we applied here adjusts for the confounding effect of both resident characteristics and “clustering” under explicit distributional assumptions. A study (Li et al. 2009) demonstrated that the mixed estimates showed close agreement with the

fixed-effects modeling when estimating another CMS QM of decline in physical function.

The performance of the mixed logistic models was measured by the c -statistic (Hanley and McNeil 1982), which measures the ability of the model to discriminate between residents with and without incontinence; the c -statistic ranges from 0.5 (random discrimination) to 1.0 (perfect discrimination).

We obtained the predicted probability of incontinence (p_{ij}) for each resident based solely on the fixed-effect components of the model (Li et al. 2009). Expected incontinence rate for each facility (E_j) was then calculated as $\sum_{j=1}^{n_j} p_{ij} / n_j$.

Finally, for each facility in each quarter, we first calculated the risk-adjusted QM as

$$\text{logit}(QM_j) = \ln\left(\frac{O_j}{1 - O_j}\right) - \ln\left(\frac{E_j}{1 - E_j}\right) + \ln\left(\frac{\bar{O}}{1 - \bar{O}}\right) \tag{1}$$

where \bar{O} is the overall incontinence rate for all residents in the quarterly sample ($\bar{O} = 47.4\%$, 47.6% , and 47.7% for the first, second, and third quarter of 2005, respectively), and then back-transformed $\text{logit}(QM_j)$ to the probability scale (Li et al. 2007):

$$QM_j = \left(1 + \text{logit}^{-1}(QM_j)\right)^{-1} \tag{2}$$

Note that we used equations (1) and (2) to calculate QM_j because a study (Li et al. 2007) has shown that it is consistent with the specification of the logistic model and better identifies facilities with extreme rankings than other measures such as those based on the difference or ratio between observed (O_j) and expected (E_j) rates. We also did not use the random intercept u_j to directly calculate the QM because u_j may be biased as a result of “statistical shrinkage” during model estimation (Li et al. 2009).

Indeed, there are both advantages and disadvantages of the shrinkage estimator u_j . First, especially for small facilities u_j is inherently biased, which would have been an issue if we used the shrinkage-based QM to define the short-term stability of facility outcome. On the other hand, the shrinkage estimator helps to avoid extreme estimation and increase the precision of the QM for small facilities, by statistically shrinking the estimate toward the grant mean. As a result, the estimates for small facilities would be more stable when computed over the short term. However, this would mask these small facilities’ real short-term fluctuations in the continence QM that we intended

to measure. We believed that the “smallness” of a facility is analogous to other facility attributes (e.g., profit status) that could impact facility performance but should not be treated as exogenous factors and adjusted for in the multivariate model. Statistically adjusting for these facility attributes would “excuse” the potential short-term instability of particular types of facilities and lead to a false impression of their real fluctuations in outcome relative to other facilities.

Impact of Case Mix on Cross-Sectional Variations of QMs

We first performed separate cross-sectional analyses on each quarterly data at the facility level. In bivariate analyses, we calculated the Pearson correlation between each QM (the CMS or risk-adjusted QM) and each of five case-mix variables. Case-mix variables included the percent of residents in the facility with age ≥ 85 years, percent of female residents, percent of residents with cognitive skill impairment, percent of residents able to make themselves understood only sometimes, percent of residents rarely able to make themselves understood, and average number of ADLs (aggregating the six ADL items described before with a range between 0 and 24).

To estimate the overall variation of the CMS or risk-adjusted QM across facilities that was explained by case mix, we further estimated multivariate least-squares models where the dependent variable was the CMS or the risk-adjusted QM, and independent variables were the case-mix variables defined above. We used the adjusted R^2 of each model to quantify the impact of case mix on the QM, that is, the higher the adjusted R^2 the more case mix can explain the cross-sectional variation of the QM. The cross-sectional analyses were performed for all NHs and for the subset of NHs with extreme CMS or risk-adjusted QM rate, that is, those at the top or bottom 25 percent rankings of the QM, or at the top or bottom 10 percent rankings.

Impact of Case Mix on the Short-Term Stability of QMs

To examine the short-term stability of QMs and the impact of case mix, we first estimated the Pearson correlation between each QM in the first quarter of 2005 and each QM in the third quarter of 2005, and the Pearson correlation of each case-mix variable between the first quarter and the third quarter of 2005.

We further defined the short-term variation (or, inversely, stability) of the QM as the difference between a facility’s QM rate in the third quarter of 2005 and its QM rate in the first quarter of 2005. Similarly, we defined the short-term variation of each case-mix variable as its difference between the

third quarter and the first quarter of 2005. Multivariate least-squares models were then estimated where the dependent variable was the short-term variation of the CMS or risk-adjusted QM, and independent variables were short-term variations of the case-mix variables. We used the adjusted R^2 to quantify how the overall short-term change in case mix can explain the short-term change in the QM.

To determine the robustness of the results to our definition of “short term,” we repeated the analyses where the longitudinal change of the QMs and case mix was calculated for a 3-month period, using data in the first and second quarters of 2005. All short-term longitudinal analyses were performed for all NHs and for the subset of NHs with least stable QM, that is, those at the top or bottom 25 percent rankings of the short-term change of QM between the third (or second) quarter and first quarter of 2005, or those at the top or bottom 10 percent rankings.

Because the QMs were calculated for facilities with different numbers of residents, all facility-level analyses (Pearson’s correlation and multivariate linear models) were weighted in reverse proportion to the number of residents (n_i) used to calculate the QM in the first quarter of 2005.

RESULTS

Descriptive Results

The data in the first quarter included 600,580 long-term residents in 10,437 facilities. Nearly half of them (47.39 percent) had urinary and/or fecal incontinence (Table 1). They were mostly females, 80 years old on average, and showed impairment in cognitive skills, behaviors, and physical function. The risk-adjustment model estimated on this dataset predicted the outcome very well, with c -statistic = 0.91.

Table 2 shows that the average rates of the CMS and the risk-adjusted QM were similar, but facilities varied considerably in both QM rates and case mix. The Kendall τ correlation between the CMS and risk-adjusted QMs was 0.45, suggesting considerable discrepancy of facility rankings derived from the two QMs. Results of data in the second and third quarters of 2005 were similar.

Table 2 also shows that the Pearson correlation for the CMS QM was 0.92 between the first and second quarter of 2005, and 0.87 between the first and third quarter of 2005; similar correlations were found for the risk-adjusted QM (0.85 and 0.78, respectively) and for most case-mix variables, suggesting high stability over the short term.

Table 2: Nursing Home Incontinence Quality Measures (QMs) and Case Mix in the First Quarter (Q1) of 2005 (n = 10,437 Nursing Homes) and Stability over the Second Quarter (Q2) and Third Quarter (Q3) of 2005

	Descriptive Statistic in Q1				Stability (Pearson's Correlation Coefficient)		
	Mean	SD	Minimum	Maximum	Between Q1 and Q2	Between Q1 and Q3	Between Q2 and Q3
CMS QM rate (%)	47.36	15.04	0	94.12	r = 0.92	r = 0.87	r = 0.87
Risk-adjusted QM rate (%)	47.12	10.88	0	92.53	r = 0.85	r = 0.78	r = 0.78
Case mix							
% of residents with age ≥ 85 years	47.90	17.93	0	94.59	r = 0.96	r = 0.95	r = 0.95
% of female residents	71.14	13.05	0	100.00	r = 0.94	r = 0.93	r = 0.93
% of residents with cognitive skill impairment	53.40	17.26	0	100.00	r = 0.93	r = 0.90	r = 0.90
% of residents only able to make self understood sometimes	11.73	9.61	0	94.12	r = 0.90	r = 0.86	r = 0.86
% of residents rarely able to make self understood	1.01	1.91	0	42.11	r = 0.76	r = 0.67	r = 0.67
Average number of ADLs	11.55	2.35	0	19.54	r = 0.96	r = 0.93	r = 0.93

ADLs, activities of daily living; CMS, Centers for Medicare and Medicaid Services; QM, quality measure.

Impact of Case Mix on the QMs

Table 3 shows that for all NHs, the CMS QM showed varied bivariate correlations with case-mix variables: the cross-sectional Pearson's correlation ranged from 0.15 for the percent of residents with cognitive skill impairment to 0.69 for the average number of ADLs; the adjusted R^2 was 0.50 in a multivariate regression that included all case-mix variables. On the other hand, the Pearson correlation coefficient of the risk-adjusted QM with each case-mix variable was <0.10 (in absolute value), and the corresponding adjusted R^2 in the multivariate model was 0.02. When analyses were limited to facilities with extreme outcome, that is, those at the top or bottom 25 percent rankings ($n = 5,220$) or at the top or bottom 10 percent rankings ($n = 2,088$) of each QM, the cross-sectional association (in both bivariate and multivariate analyses) between each QM and case mix tended to increase. Nonetheless, the association of the risk-adjusted QM with case mix continued to be substantially lower compared with the CMS QM.

Table 4 presents the short-term stability of the QMs and the relationship between the short-term change of the QMs and the short-term change in case mix. The risk-adjusted QM was slightly less stable than the CMS QM. For example, the correlation between the first and third quarter of 2005 was 0.78 for the risk-adjusted QM and 0.87 for the CMS QM for all NHs. Similar to results in the cross-sectional analyses, the short-term variation of the CMS QM was partially explained by case mix: the adjusted R^2 in the multivariate regression of the CMS QM was 0.26 for all facilities, 0.38 for facilities at the top or bottom 25 percent rankings of short-term variation, and 0.48 for those at the top or bottom 10 percent rankings. On the other hand, the impact of case mix on the risk-adjusted QM was substantially reduced, with adjusted R^2 ranging from 0.01 to 0.03 in alternative cases. Similar results were found when the "short term" was defined as 3 months rather than 6 months (results not shown).

DISCUSSION

The NH industry has seen an increasing use of report cards for market-driven quality improvement (GAO 2002; Abt 2006; Mukamel et al. 2007). Given the intended use of the QMs, it is critical that they reflect the true performance variations rather than case-mix variations across facilities.

This study examined the impact of case mix on two aspects of the CMS incontinence QM—its cross-sectional variation and longitudinal stability in

Table 3: Cross-Sectional Correlation between Nursing Home Quality Measure (QM) of Incontinence and Case Mix: First Quarter of 2005

	CMS QM			Risk-Adjusted QM		
	All Nursing Homes (n = 10,437)	Top 25% and Bottom 25% Nursing Homes (n = 5,220)	Top 10% and Bottom 10% Nursing Homes (n = 2,088)	All Nursing Homes (n = 10,437)	Top 25% and Bottom 25% Nursing Homes (n = 5,220)	Top 10% and Bottom 10% Nursing Homes (n = 2,088)
% of residents with age ≥ 85	$r = 0.34$	$r = 0.43$	$r = 0.55$	$r = 0.04$	$r = 0.07$	$r = 0.13$
% of female residents	$r = 0.32$	$r = 0.41$	$r = 0.53$	$r = 0.02$	$r = 0.05$	$r = 0.11$
% of residents with cognitive skill impairment	$r = 0.15$	$r = 0.18$	$r = 0.17$	$r = -0.08$	$r = -0.10$	$r = -0.14$
% of residents only able to make self understood	$r = 0.29$	$r = 0.37$	$r = 0.47$	$r = 0.02$	$r = 0.03$	$r = 0.07$
% of residents rarely able to make self understood	$r = 0.18$	$r = 0.22$	$r = 0.25$	$r = 0.04$	$r = 0.06$	$r = 0.08$
Average number of ADLs	$r = 0.69$	$r = 0.77$	$r = 0.84$	$r = -0.08$	$r = -0.09$	$r = -0.06$
Adjusted R^2	0.50	0.62	0.71	0.02	0.04	0.07

Note. Top and bottom nursing homes are defined based on rankings of the CMS or risk-adjusted QM rate, respectively.

*Derived from the multivariate ordinary least-squares model of the QM as a function of all case-mix variables.

ADL, activities of daily living; CMS, Centers for Medicare and Medicaid Services; QM, quality measure.

Table 4: Stability of the Nursing Home Incontinence Quality Measure (QM) between the First Quarter (Q1) and Third Quarter (Q3) of 2005 and the Impact of Case Mix

	<i>CMS QM</i>		<i>Risk-Adjusted QM</i>	
	<i>Stability*</i>	<i>Adjusted R²</i>	<i>Stability*</i>	<i>Adjusted R²</i>
All nursing homes (<i>n</i> = 9,799)	<i>r</i> = 0.87	0.26	<i>r</i> = 0.78	0.01
Top 25% and bottom 25% nursing homes (<i>n</i> = 4,900)	<i>r</i> = 0.72	0.38	<i>r</i> = 0.60	0.01
Top 10% and bottom 10% nursing homes (<i>n</i> = 1,960)	<i>r</i> = 0.49	0.48	<i>r</i> = 0.36	0.03

Note. Top and bottom nursing homes are defined based on rankings of the short-term change of the QM (CMS or risk-adjusted QM) between Q1 and Q3 of 2005. The stabilities for all nursing homes in the CMS QM (*r* = 0.87) and the risk-adjusted QM (*r* = 0.78) were identical to the ones shown in Table 2.

*Pearson’s correlation of the QM between Q1 and Q3 of 2005.

†Derived from the multivariate ordinary least-squares model of short-term change of the QM as a function of short-term change of case mix. See Table 2 for definitions of case-mix variables. Short-term change of QM (or case mix) is defined as the QM (or case mix) in Q3 minus the QM (or case mix) in Q1 of 2005.

CMS, Centers for Medicare and Medicaid Services; QM, quality measure.

the short run. Both properties are important because they together determine how the QM can be used to support valid outcome comparisons. Our findings are consistent with the hypotheses regarding the impact of case mix on the QM. We first demonstrated that at least half of the cross-sectional (or between-facility) variation of the CMS QM was explained by facility case mix (Table 3). In addition, both the CMS QM and case mix showed relatively high stability over the short term and as a result of minimal risk adjustment of the QM, over 25 percent of its short-term variation was explained by case mix. These case-mix impacts were more substantial for facilities with more extreme or unstable CMS QM rates. We developed a risk-adjusted QM and found that it was much less susceptible to the cross-sectional and short-term longitudinal impacts of case mix, even for the subset of facilities with more extreme or unstable rates.

We defined case mix based on resident demographics as well as cognitive, behavioral, and physical functions, which affect residents’ risk of being incontinent (Ouslander and Schnelle 1995) but are generally beyond the control of facilities’ practices in continence care. For example, residents with a higher level of physical limitations are more likely to be incontinent than other residents even under the same quality of continence care. Therefore, if the

intended use of the QM is to measure continence care quality, the QM should be adjusted for exogenous risk factors. Given the multidimensional nature of NH care, physical function and other risk components of incontinence may also reflect other aspects of care practice. Nonetheless, this study focuses on the incontinence outcome and we do not feel that one QM could be appropriately used for profiling multiple aspects of NH care. The issue of multidimensionality of NH quality and its implications for risk adjustment were discussed before (Mukamel et al. 2008a).

Studies have compared the CMS QMs and further risk-adjusted QMs in ranking facilities and reported that cross-sectional rankings tended to disagree (Arling et al. 2007; Mukamel et al. 2008a; Li et al. 2009). These studies conclude that further risk-adjusted QMs may be an improvement over existing QMs, assuming that more extensive risk adjustment can successfully remove the influence of case mix on outcomes. Our study explicitly tested this assumption and confirmed that the overall case-mix impact can be minimized. However, when we focused on facilities with more extreme outcome rankings, the risk-adjusted incontinence QM still correlated, although to a much lesser extent, with several case-mix variables (Table 3). This raises an issue related to current NH quality improvement efforts, which tend to target facilities with extreme performance. For example, state surveyors may exert more extensive oversight on facilities with poorer outcomes, or P4P programs may be designed to financially reward facilities with the best outcomes. We recommend caution in dealing with these “extreme” facilities because their QM rate, even with extensive risk adjustment, may still be affected by case mix.

Our longitudinal analyses found that although the CMS incontinence QM and case mix in general showed a high level of stability (Table 2), the risk-adjusted QM tended to show lower stability over the short term than the CMS QM (Tables 2 and 4). This can be explained by the high stability of case mix and the fact that the risk-adjusted QM has adjusted away the confounding effect of case mix. Therefore, the slightly reduced short-term stability of the risk-adjusted QM should not be interpreted as it being less useful in guiding outcome comparisons. Rather, the risk-adjusted QM, when successfully removing the longitudinal impact of case mix, can more accurately reflect a facility’s possible performance fluctuations over the short term.

We note that a subset of facilities may show especially high fluctuation in outcomes as indicated by the high short-term variation of both the CMS and risk-adjusted QMs (Table 4: $r < 0.5$ for the subset of 1960 facilities). Because this fluctuation also manifests on the extensively risk-adjusted QM, it likely reflects the volatility of these facilities’ actual continence care performance.

Our results suggest two important policy implications from a methodological perspective: (1) analysts should explicitly test, rather than assume, that statistical risk adjustment removes the confounding effect (both cross-sectional and longitudinal) of case mix on outcome variations; and (2) the cross-sectional and short-term longitudinal impacts of case mix may be more substantial for facilities with more extreme or unstable outcomes.

One of the limitations of this study is that we only examined one CMS QM, and our findings may not be generalizable to other QMs. Another limitation is that we cannot determine the reasons for the short-term volatility of the subset of facilities' incontinence outcome shown in Table 4. Berlowitz et al. (1998), when focusing on pressure ulcer rates in a group of VA NHs, have suggested that chance effects may play a role in the short-term stability of the outcome, and that small facilities tend to show lower stability. However, the outcome examined in this study, namely incontinence, is much more frequent (47 percent) than the one in Berlowitz's study (3 percent on average). Therefore, chance effects may not play an important role in interpreting the high volatility of incontinence rate among the subset of 1960 facilities in our study. In addition, these 1960 facilities have an average number of 51 residents (compared with 59 residents for all facilities), and all facilities in our study have at least 30 chronic care cases according to CMS's definition (Abt 2004). Thus, facility size may not be a major reason either for the results we observed in these facilities.

A potential reason for these facilities' high volatility in incontinence rate is abrupt staff turnover during the study period, which may change practice and outcome even in the short term. In addition, acute illness such as respiratory or gastrointestinal infection in NHs tends to come in clusters because of contagion. When the infection is transmitted throughout a facility, it can affect the continence outcome for a group of patients, and thus cause abrupt change of the facility's incontinence rate in the short term. Further studies are needed to examine these and other potential reasons for the results we found.

In conclusion, our study focuses on the CMS incontinence QM and tests how facility case mix may affect its cross-sectional variation and short-term stability. The findings suggest considerable impacts of case mix on the two important properties of the QM. Further risk adjustment on this QM can minimize the overall impact of case mix and make the QM more likely to reflect a facility's cross-sectional standing and short-term stability in continence care practice. Using the risk-adjusted QM for inference of performance, however, still needs to be made with caution because a number of NHs in the nation could be volatile in actual care practice over the short term.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded by the National Institute on Aging under grant AG027420. An earlier version of this work was presented at the Academy Health Annual Research Meeting in June 2009 (Chicago, IL) and at the 137th APHA Annual Meeting in November 2009 (Philadelphia, PA).

Disclosures: None.

Disclaimers: None.

REFERENCES

- Abt. 2004. *Quality Measures for National Public Reporting: User's Manual, v1.2*. Cambridge, MA: Abt Associates Inc.
- . 2006. *Quality Monitoring for Medicare Global Payment Demonstrations: Nursing Home Quality-Based Purchasing Demonstration*. Cambridge, MA: Abt Associates Inc.
- AHRQ. 1996. *U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. Urinary Incontinence in Adults: Acute and Chronic Management. Clinical Practice Guideline. Number 2, 1996 Update*. Rockville, MD: AHRQ.
- Arling, G., T. Lewis, R. L. Kane, C. Mueller, and S. Flood. 2007. "Improving Quality Assessment through Multilevel Modeling: The Case of Nursing Home Compare." *Health Services Research* 42 (3, Part 1): 1177–99.
- Berlowitz, D. R., J. J. Anderson, A. S. Ash, G. H. Brandeis, H. K. Brand, and M. A. Moskowitz. 1998. "Reducing Random Variation in Reported Rates of Pressure Ulcer Development." *Medical Care* 36 (6): 818–25.
- Chassagne, P., I. Landrin, C. Neveu, P. Czernichow, M. Bouaniche, J. Doucet, P. Denis, and E. Bercoff. 1999. "Fecal Incontinence in the Institutionalized Elderly: Incidence, Risk Factors, and Prognosis." *American Journal of Medicine* 106 (2): 185–90.
- Elliott, M. N., R. Swartz, J. Adams, K. L. Spritzer, and R. D. Hays. 2001. "Case-Mix Adjustment of the National CAHPS Benchmarking Data 1.0: A Violation of Model Assumptions?" *Health Services Research* 36 (3): 555–73.
- General Accounting Office (GAO). 2002. *Public Reporting of Quality Indicators Has Merit, but National Implementation Is Premature (Publication No. GAO-03-187)*. Washington, DC: GAO.
- . 2007. *Nursing Homes: Federal Actions Needed to Improve Targeting and Evaluation of Assistance by Quality Improvement Organizations, May 2007*. Washington, DC: GAO.
- Green, J., L. J. Passman, and N. Wintfeld. 1991. "Analyzing Hospital Mortality. The Consequences of Diversity in Patient Mix." *Journal of the American Medical Association* 265 (14): 1849–53.
- Hanley, J. A., and B. J. McNeil. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143 (1): 29–36.

- Harrington, C., J. T. Mullan, and H. Carrillo. 2004. "State Nursing Home Enforcement Systems." *Journal of Health Politics, Policy and Law* 29 (1): 43–73.
- Hawes, C., J. N. Morris, C. D. Phillips, V. Mor, B. E. Fries, and S. Nonemaker. 1995. "Reliability Estimates for the Minimum Data Set for Nursing Home Resident Assessment and Care Screening (MDS)." *Gerontologist* 35 (2): 172–8.
- Iezzoni, L. I. 2003. *Risk Adjustment for Measuring Health Care Outcomes*. Chicago, IL: Health Administration Press.
- Institute of Medicine. 2001. *Improving the Quality of Care in Nursing Homes*. Washington, DC: National Academies Press.
- Karon, S. L., F. Sainfort, and D. R. Zimmerman. 1999. "Stability of Nursing Home Quality Indicators over Time." *Medical Care* 37 (6): 570–9.
- Lawton, M. P., R. Casten, P. A. Parmelee, K. Van Haitsma, J. Corn, and M. H. Kleban. 1998. "Psychometric Characteristics of the Minimum Data Set II: Validity." *Journal of the American Geriatrics Society* 46 (6): 736–44.
- Levy-Storms, L., J. F. Schnelle, and S. F. Simmons. 2007. "What Do Family Members Notice Following an Intervention to Improve Mobility and Incontinence Care for Nursing Home Residents? An Analysis of Open-Ended Comments." *Gerontologist* 47 (1): 14–20.
- Li, Y., X. Cai, L. G. Glance, W. D. Spector, and D. B. Mukamel. 2009. "National Release of the Nursing Home Quality Report Cards: Implications of Statistical Methodology for Risk Adjustment." *Health Services Research* 44 (1): 79–102.
- Li, Y., A. W. Dick, L. G. Glance, X. Cai, and D. B. Mukamel. 2007. "Misspecification Issues in Risk Adjustment and Construction of Outcome-Based Quality Indicators." *Health Services and Outcomes Research Methodology* 7 (1–2): 39–56.
- Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberger. 2006. *SAS for Mixed Models*. 2d Edition. Cary, NC: SAS Institute Inc.
- Mor, V., J. Angelelli, R. Jones, J. Roy, T. Moore, and J. Morris. 2003. "Inter-Rater Reliability of Nursing Home Quality Indicators in the U.S." *BMC Health Services Research* 3 (1): 20–33.
- Mukamel, D. B., L. G. Glance, Y. Li, D. L. Weimer, W. D. Spector, J. S. Zinn, and L. Mosqueda. 2008a. "Does Risk Adjustment of the CMS Quality Measures for Nursing Homes Matter?" *Medical Care* 46 (5): 532–41.
- Mukamel, D. B., W. D. Spector, J. S. Zinn, L. Huang, D. L. Weimer, and A. Dozier. 2007. "Nursing Homes' Response to the Nursing Home Compare Report Card." *Journal of Gerontology Series B Psychological Sciences and Social Sciences* 62 (4): S218–25.
- Mukamel, D. B., N. M. Watson, H. Meng, and W. D. Spector. 2003. "Development of a Risk-Adjusted Urinary Incontinence Outcome Measure of Quality for Nursing Homes." *Medical Care* 41 (4): 467–78.
- Mukamel, D. B., D. L. Weimer, W. D. Spector, H. Ladd, and J. S. Zinn. 2008b. "Publication of Quality Report Cards and Trends in Reported Quality Measures in Nursing Homes." *Health Services Research* 43 (4): 1244–62.
- Nelson, R., S. Furner, and V. Jesudason. 1998. "Fecal Incontinence in Wisconsin Nursing Homes: Prevalence and Associations." *Diseases of the Colon and Rectum* 41 (10): 1226–9.

- Nelson, R. L., and S. E. Furner. 2005. "Risk Factors for the Development of Fecal and Urinary Incontinence in Wisconsin Nursing Home Residents." *Maturitas* 52 (1): 26–31.
- O'Malley, A. J., A. M. Zaslavsky, M. N. Elliott, L. Zaborski, and P. D. Cleary. 2005. "Case-Mix Adjustment of the CAHPS Hospital Survey." *Health Services Research* 40 (6, Part 2): 2162–81.
- Ouslander, J. G., and J. F. Schnelle. 1995. "Incontinence in the Nursing Home." *Annals of Internal Medicine* 122 (6): 438–49.
- Rantz, M. J., L. Hicks, G. F. Petroski, R. W. Madsen, D. R. Mehr, V. Conn, M. Zwiygart-Staffacher, and M. Maas. 2004. "Stability and Sensitivity of Nursing Home Quality Indicators." *Journal of Gerontology Series A Biological Sciences and Medical Sciences* 59 (1): 79–82.
- Schnelle, J. F., C. A. Alessi, S. F. Simmons, N. R. Al-Samarrai, J. C. Beck, and J. G. Ouslander. 2002. "Translating Clinical Research into Practice: A Randomized Controlled Trial of Exercise and Incontinence Care with Nursing Home Residents." *Journal of the American Geriatrics Society* 50 (9): 1476–83.
- Schnelle, J. F., and F. W. Leung. 2004. "Urinary and Fecal Incontinence in Nursing Homes." *Gastroenterology* 126 (1, Suppl 1): S41–7.
- Schnelle, J. F., P. McNees, V. Crooks, and J. G. Ouslander. 1995. "The Use of a Computer-Based Model to Implement an Incontinence Management Program." *Gerontologist* 35 (5): 656–65.
- Schnelle, J. F., D. Newman, M. White, J. Abbey, K. A. Wallston, T. Fogarty, and M. G. Ory. 1993. "Maintaining Continence in Nursing Home Residents through the Application of Industrial Quality Control." *Gerontologist* 33 (1): 114–21.
- Simmons, S. F., E. T. Garcia, M. P. Cadogan, N. R. Al-Samarrai, L. F. Levy-Storms, D. Osterweil, and J. F. Schnelle. 2003. "The Minimum Data Set Weight-Loss Quality Indicator: Does It Reflect Differences in Care Processes Related to Weight Loss?" *Journal of the American Geriatrics Society* 51 (10): 1410–8.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2. The Initial List of Potential MDS Predictors of the CMS Incontinence QM.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.