Resident, Nursing Home, and State Factors Affecting the Reliability of Minimum Data Set Quality Measures

Ning Wu, PhD
Vincent Mor, PhD
Jason Roy, PhD

Nursing home quality measures impact policy decisions such as reimbursement or consumer choice. Quality indicators in the United States are collected through the federally mandated Minimum Data Set (MDS). Bias in MDS data collection or coding can thus have a negative impact on policy applications. To understand whether bias was present in coding, the authors studied 5174 pairs of MDS assessments that were independently collected by nursing home staff and study nurses from 206 nursing homes. The authors developed multivariate multilevel models to identify nursing home and resident characteristics that were significantly associated with the data quality of multiple MDS measures of nursing home quality. The outcomes were coding differences between nursing home staff and study nurses. Resident characteristics explained little of the variation in coding differences among facilities, while facilities characteristics explained 4% to 20% of the variation and state location further explained 13% to 34% of the variation. A generalized effect of nursing home state location tended to be consistent across measures. States that overidentified problems also tended to have worse quality indicators and vice versa. Comparisons of MDS-based quality indicators reflect differences in assessment practices at least as much as true quality differences. Efforts to standardize assessment practices across states are needed. (Am J Med Qual 2009;24:229-240)

Keywords: Minimum Data Set; measurement bias; quality indicator; multilevel modeling; nursing home

The Minimum Data Set (MDS), data collected during comprehensive nursing home resident assessments, contains information on residents' medical, social, and functional status. The assessment data were meant to assist facility staff to detect symptoms and signs in a timely manner and to develop appropriate care plans.[1-3] Since the 1990s, the federal government of the United States has mandated that nursing homes collect and submit the MDS on Medicare and/or Medicaid beneficiaries who reside in these facilities. There are instruction manuals covering completion of the assessment process, and federal and state officials as well as state provider associations, engage in routine training.

Uniform data collection also permits monitoring and public reporting of provider quality, and forms the basis for payment.[4] Quality indicators (QIs) were constructed and calculated to reflect the quality of care in nursing homes. The QIs are facility aggregates of selected sets of individual MDS items. They represent prevalence or incidence estimates of clinical events and, as such, are said to characterize the quality of care in individual facilities. Currently, the US government publishes MDS-based QIs online to: (*a*) assist consumers to make informed decisions when choosing a provider, and (*b*) stimulate providers to compete on the basis of quality.[5] In addition, federal and some state governments used the MDS to develop case mix-based reimbursement systems, which determine payment for nursing home care.[6,7]

AUTHORS' NOTE: Dr Wu is with Abt Bio-pharma Solutions, Inc, Lexington, MA; Dr Mor is with the Department of Community Health, Providence, RI; and Dr Roy is with the Geisinger Center for Health, Danville, PA. The authors have no conflicts of interest to disclose. *Corresponding author:* Ning Wu, PhD, Abt Bio-pharma Solutions, Inc, 181 Spring Street. Lexington, MA 02421 (email: ning.wu@abtbiopharma.com).

American Journal of Medical Quality, Vol. 24, No. 3, May/June 2009
DOI: 10.1177/1062860609332510

Proper applications of the data are contingent on the MDS data being consistent and reliable. Although a number of studies suggested that pooled MDS data collected from multiple nursing homes were valid and reliable,[8-11] the data quality varied among nursing homes.[12-16] If differential assessment practices were large, the use of the assessment data for policy applications would be compromised. At a minimum, the QIs derived from the MDS data would not be comparable and might mislead consumers' choices as well as providers' own appreciation of their performance relative to their peers.

Researchers hypothesized that certain nursing home characteristics (eg, staff training, nursing home specialty) may be associated with MDS quality.[10,12,13,15,17] However, few existing studies have found empirical evidence to support this hypothesis.[17] In fact, most studies of MDS quality used data from a small number of nursing homes and were not powered to detect such associations. The largest study, by Mor et al, was based on more than 5000 pairs of MDS assessments from 209 nursing homes. Nevertheless, no facility characteristics were found to be significantly associated with the reliability of MDS data.[13] One probable explanation is that the reliability of MDS measures was quantified by kappa, a chance-corrected measure of agreement. Kappa does not measure the direction of measurement errors. Thus, if some nursing homes had more overcoding errors than other nursing homes (or vice versa), comparison of kappa would not detect such differences.

Furthermore, existing studies of MDS quality either focused on 1 MDS item or treated measurements in multiple health domains as uncorrelated. We reasoned that facility-specific data collection behaviors might affect the quality of assessment data in multiple health domains. For example, a nursing home that did a poor job in detecting incontinence may also underdocument pressure ulcers. If such patterns exist, the cumulative effect of measurement errors may be substantial on composite measures that summarize nursing home quality of care in multiple health domains.

METHOD

We studied the association between measurement errors in 8 MDS-derived measurement scales and the characteristics of residents and nursing homes. We analyzed data of 5344 pairs of MDS assessments that were independently collected by facility, staff, and research nurses in 206 nursing homes. Measurement errors were defined as the difference in MDS coding between nursing home staff and study nurses. Multivariate multilevel models were built to identify the factors that were strongly associated with measurement errors, tested whether facility staff consistently tend to have biased measures in multiple health domains, and explored the association between measurement biases in MDS scales and the MDS-derived QIs.

Data Source

We analyzed the data collected in the National Study to Validate the Long-Term and Postacute Care Quality Indicators Derived from MDS (referred to as the QI validation study). The purpose of the study was to validate a set of QIs of freestanding or hospital-based nursing homes for quality of care monitoring and public reporting. Detailed information about the study design and data collection procedure is published elsewhere.[10] In this subsection, we briefly describe the data collection procedures that are relevant to our study.

In the QI validation study, a research team collected reliability data in 209 facilities in 6 states (California, Illinois, Missouri, Ohio, Pennsylvania, and Tennessee) during 2001 and 2002. The participating facilities were among 462 facilities that were randomly selected and approached by the study team. The participating and nonparticipating nursing homes were similar, except for the nonparticipating group being slightly less likely to be in a chain or for profit.[13]

In each participating nursing home, study nurses selected up to 30 residents who had a recently completed MDS and conducted independent MDS assessments. The average interval between the paired assessments was 23 days (standard deviation, 19 days). When completing the MDS, both study nurses and facility staff were expected to follow the assessment protocol contained in the *Resident Assessment Instrument (RAI) User's Manual*.[10] No additional assessment instructions were given to the study nurses.

Several measures were taken to ensure the integrity of data collected by study nurses. All but

1 of the study nurses involved in the study were registered nurses with long-term care experiences. Study nurses underwent a 5-day training prior to the site visits and had weekly teleconferences to resolve any problems they may have encountered during the data collection phase. Data were not available on the type or amount of MDS training received by nursing home staff. Completion of MDS requires that nurses use all sources of information including assessing residents, interviewing nursing home staff and attending physicians, and reviewing medical charts.[10] To reduce data contamination, study nurses were prohibited from reviewing the facility MDS record or conducting other data collection activities before completing their own assessment of residents. Study nurses also completed up to 2 pairs of assessments per facility and their reliability was assessed. Interrater reliability was high (kappa > 0.85) among study nurses on virtually all assessment items.[10,13]

We merged the reliability data with the 2000 to 2002 MDS national repository file to determine residents' demographic characteristics and clinical conditions. Information on nursing home characteristics was obtained from the 2001 Online Survey Certification and Reporting (OSCAR) data. We received a data use agreement from the Centers for Medicare and Medicaid Services (CMS) and the Brown University Institutional Review Board approved this project.

Analytical Approach

Our analyses were based on 5174 pairs of MDS assessments collected from 206 facilities. We excluded 150 residents from the analytic sample whose assessment data were missing. Compared to the remaining residents, the excluded residents were less likely to be white and more likely to live in nursing homes in California or Ohio. To obtain reliable estimates for the effect of facility characteristics on data quality, we further excluded 3 nursing homes that had less than 10 eligible participating residents.

Our analyses focused on the measurement errors of 8 MDS-derived ordinal scales. The MDS version 2.0 contains over 350 items that document the residents' demographics, diagnoses and treatment, and physical and mental functioning. The selected scales were derived from 29 MDS items that measured 7 important health domains. Many of the comprising MDS items are included in the nursing home QI estimations that CMS releases to the public. Some of the selected health domains were also prone to measurement errors in the nursing home population.[16,18-20] The 8 scales are as follows: (1) activities of daily living (ADL), (2) Cognitive Performance Scale (CPS), (3) pain, (4) restraints, (5) bowel incontinence, (6) bladder incontinence, (7) pressure ulcer (PU) staging, and (8) delirium. All of the scales, except for CPS, were composed of MDS items in the same health domain. More details about the scales and the MDS items are provided in the Appendix.

We first compared the proportion of MDS records with overcoding and undercoding errors among facilities. Previous work on this data file suggested that study nurses' coding was consistently reliable.[13] Thus, we treated the study nurses' coding as a comparison standard and defined overcoding as the situation of facility ratings being higher than the ratings of the study nurses and undercoding as the situation of facility ratings being lower than the ratings of the study nurses.

We then built a 2-level multivariate multilevel model to identify factors that were associated with measurement errors in 1 or multiple measures. Measurement errors were included as the outcomes and treated as a continuous property for all scales, with a larger difference between nursing home and study nurses' ratings indicating more interrater disagreement. Resident and nursing home characteristics were included as explanatory variables to estimate their impact on the direction and magnitude of measurement errors.

We started from a null model by including facility-level random intercepts as the explanatory variable, indicated in the equation below. The purpose was to quantify the within-nursing home and between-nursing home variation in the measurement errors. The null model can be expressed as

$$y_{ij}^{(k)} = \beta_{0j}^{(k)} + \varepsilon_{ij}^{(k)}$$

$$\beta_{0j}^{(k)} = \beta_{0}^{(k)} + u_{0j}^{(k)}$$

where k indices the different MDS scales, i indices the ith resident, j indicates the jth facility, and $\beta_{0j}^{(k)}$ indices the random intercept for the kth outcome of the jth facility. A total of 8 sets of regression coefficients were estimated simultaneously for the outcomes. The random intercepts

($\beta_{0j}^{(k)}$) were assumed to be independent and normally distributed, with a mean of 0 and unstructured 8 × 8 covariance matrix. A significant covariance estimate suggests that facilities that overcoded one measurement also overcoded (if the covariance is positive) or undercoded (if the covariance is negative) the other measures. Similarly, the resident-level error terms for MDS scale k ($\varepsilon_{0j}^{(k)}$) were assumed to be independent and normally distributed with a mean of zero. An unstructured 8 × 8 covariance matrix was used to describe within-resident correlation between the MDS scales. The maximum likelihood method was used for estimation.

Using the variance and covariance estimates from the null model as the referent, we sequentially added to the model resident characteristics, nursing home characteristics, and indicator variables of the state in which the nursing homes were located. We then observed the reduction in the unexplained variation of measurement errors. The full model can be expressed as

$$y_{ij}^{(k)} = \beta_{0j}^{(k)} + \sum_n \beta_n^{(k)} \text{ resident}_{ijn} + \varepsilon_{ij}^{(k)}$$

$$\beta_{0j}^{(k)} = \beta_0^{(k)} + \sum_m \beta_m^{(k)} \text{ facility}_{jm} + u_{0j}^{(k)}$$

Everything else being the same as the null model, the full model included n resident characteristics and m facility characteristics as the explanatory variables. Factors that were consistently associated with multiple outcomes were identified from the full model.

We further examined the association between measurement biases in MDS scales and the MDS-derived QIs. Because the QIs are derived from MDS items that composed the MDS scales, overcoding of the scales may bias the QIs upward, and vice versa. To test this possibility, we investigated the pattern between the state averages of QIs and predicted averages of measurement bias with consideration of observed measurement bias and adjustments for differences in resident case mix and facility characteristics. The predicted averages of measurement bias per state were derived from the fitted models using information on resident characteristics, facility characteristics, and location of the facilities. With all of the covariates set at their means and varying values for state indicator variables, we used the corresponding set of estimated regression coefficients to calculate the predicted values for each outcome (ie, the difference between the facility and research nurse raters' assessment scores on the relevant items).

Table 1

Description of Participating Residents and Nursing Homes

Variables	Count	%
Resident characteristics (n = 5174)		
Female	3638	70.3
White	4306	83.2
Age ≥ 85	2016	39.0
Dementia	1538	29.7
Nursing home characteristics (n = 206)		
Run for profit	110	53.4
Hospital based	59	28.6
Have Alzheimer's units	36	17.5
Number of deficiencies >50th percentile in state	81	39.3
Part of a chain	112	54.4
Total registered nurse hours per patient day > nation's 75th percentile in 2001 (0.49)	104	50.5
Total admin nurse full-time employee > nation's 75th percentile in 2001 (3.86)	69	33.5
Residents/on-staff nurse aide FTE > nation's 75th percentile in 2001 (3.04)	64	31.1
Number of participating nursing homes per state		
California	31	15.1
Illinois	39	18.9
Missouri	26	12.6
Ohio	33	16.0
Pennsylvania	45	21.8
Tennessee	32	15.5

RESULTS

The characteristics of our study sample matched the general descriptions of the US nursing home population. The majority of participating residents were female, white, and 39% were 85 years of age and older. About one third had dementia (Table 1). Compared to nursing homes nationwide, a higher proportion of participating nursing homes were hospital-based (29% vs 12%), a slightly lower proportion were for-profit (53% vs 65%), and a similar proportion

Table 2

Average Proportion of Facility Minimum Data Set with Undercoding or Overcoding Errors

Facility Rating Relative to Study Nurses' Rating[a]	All Participating NHs, % (n = 206)		Average Proportion by NH Location, %					
	Mean	Range	CA	IL	MO	OH	PA	TN
Activities of daily living								
Undercoding	27	(0–78)	18	36	36	21	25	26
Same	29	(3–100)	36	31	25	17	29	37
Overcoding	44	(0–91)	46	33	39	62	46	37
Cognitive Performance Scale								
Undercoding	23	(0–56)	17	26	28	18	26	25
Same	61	(9–100)	71	64	55	49	63	64
Overcoding	15	(0–61)	12	11	17	32	11	11
Delirium								
Undercoding	23	(0–73)	8	16	41	26	27	20
Same	66	(15–100)	80	74	47	57	62	70
Overcoding	12	(0–73)	11	10	12	17	10	10
Bowel incontinence								
Undercoding	13	(0–77)	8	10	25	17	13	10
Same	78	(18–100)	87	86	67	74	72	85
Overcoding	8	(0–50)	6	4	8	9	15	5
Bladder incontinence								
Undercoding	15	(0–39)	9	14	24	17	16	12
Same	76	(52–100)	83	79	66	72	74	81
Overcoding	9	(0–36)	9	7	10	11	10	7
Pressure ulcer staging								
Undercoding	9	(0–75)	5	7	12	10	10	11
Same	86	(21–100)	87	88	84	84	85	85
Overcoding	5	(0–48)	8	5	3	6	5	4
Pain								
Undercoding	19	(0–68)	11	19	30	18	20	19
Same	63	(28–100)	68	66	54	54	63	69
Overcoding	18	(0–37)	21	15	16	28	17	12
Restraints								
Undercoding	7	(0–55)	7	4	9	3	14	6
Same	90	(45–100)	90	95	89	94	84	92
Overcoding	3	(0–27)	4	1	2	4	2	2

[a]We treated the study nurses' rating as a comparison standard and defined overcoding (undercoding) as facility ratings being higher (lower) than the ratings given by study nurses. NH = nursing home; CA = California; IL = Illinois; MO = Missouri; OH = Ohio; PA = Pennsylvania; TN = Tennessee.

belonged to a chain (54% vs 56%). The staffing level of participating nursing homes was higher on average than nursing homes in the nation. Half of the participating facilities had total registered nurse hours per resident day greater than the nation's 75th percentile (0.49).

On average, nursing home staff were more likely to undercode rather than overcode the 8 scales except for ADL (Table 2). There was a substantial variation in the proportion of records with undercoding and overcoding errors between nursing homes. For example, staff in some nursing homes overcoded ADL for almost all the selected residents and the opposite was true in some other nursing homes. The MDS quality also varied by state. Nursing homes in Ohio had the highest proportion of overcoding in 6 out of the 8 scales, whereas nursing homes in Missouri had the highest proportion of undercoding in 7 scales.

The variances of all 8 outcomes estimated from the null model were significant at both facility level and resident level, suggesting that for each MDS scale the rating differences between nursing home staff and study nurses varied substantially within and between facilities. Several covariance terms at the facility level were also significant (Table 3), implying that on average, facility staff

Table 3

Variance Covariance Matrix at Facility Level Estimated From the Null Model With Random Intercepts as the Explanatory Variable[a]

Variances/ Covariance	Activities of Daily Living	Pain	Cognitive Performance Scale	Delirium	Pressure Ulcer Staging	Bowel Incontinence	Bladder Incontinence	Restraints
Activities of daily living	2.006[c]							
Pain	0.276[c]	0.162[c]						
Cognitive Performance Scale	0.218[c]	0.060[c]	0.161[c]					
Delirium	0.098	0.052[b]	0.062[c]	0.187[c]				
Pressure ulcer staging	–0.009	0.010	0.016[b]	0.015[b]	0.016[c]			
Bowel incontinence	0.051	0.029[c]	0.012	0.022[b]	0.005	0.030[c]		
Bladder incontinence	0.050	0.026[c]	0.024[c]	0.037[c]	0.006[b]	0.015[c]	0.018[c]	
Restraints	0.062[b]	0.014	0.012	0.024[b]	0.005	0.003	0.001	0.034[c]

[a]Results are from the null multilevel multivariate models with random intercepts as the explanatory variables. The outcomes are the coding differences between nursing home staff and study nurses on 8 MDS scales. The values in the diagonal cells are the facility-level variances, and the values in the off-diagonal cells are the facility level covariance between 2 measurement scales. A significant covariance estimate suggests that facilities that overcoded 1 measurement also overcoded (if the covariance is positive) or undercoded (if the covariance is negative) the other measures.
[b]The estimated variance or covariance was significant at $P < .05$.
[c]The estimated variance or covariance was significant at $P < .01$.

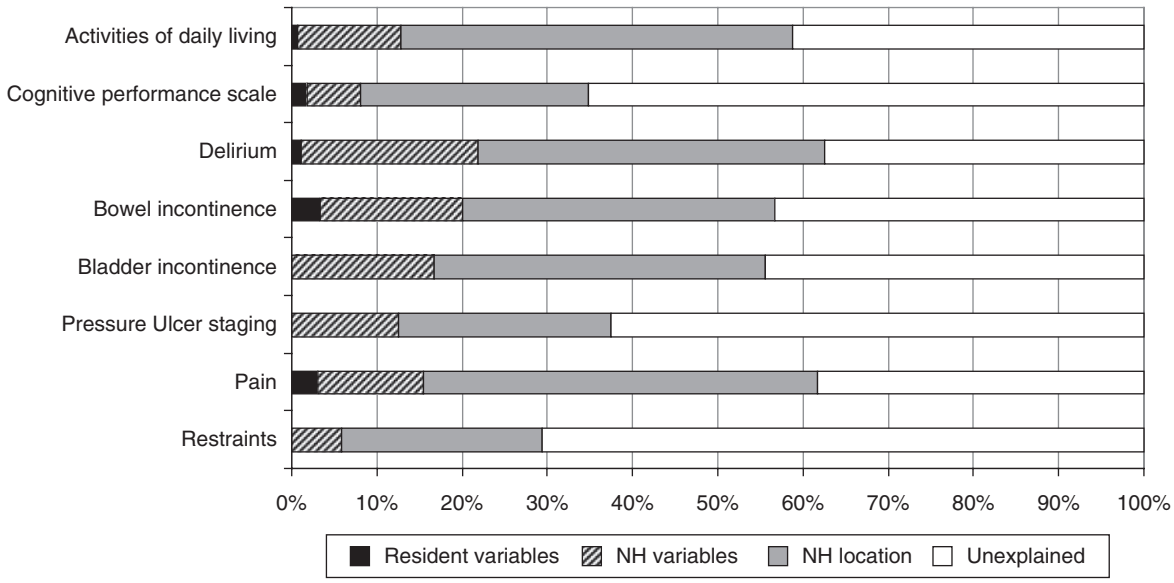

Figure 1. Percentage of nursing home level variances in the measurement errors explained by different characteristics of residents and nursing homes. Using the estimates of nursing home level variance from the null model as the referent, the authors observed the reduction in the unexplained variance by sequentially adding resident characteristics, nursing home characteristics, and nursing home location to the null model. With resident and facility variables in the model, the inclusion of nursing home location further explained the facility level variances by 13 to 34 percentage points. NH = nursing home.

who overrated residents' pain relative to study nurses also overrated cognitive impairment, and other outcomes. Other associations found include measurement bias of delirium being positively correlated with that of PU staging, incontinence, and restraints, and measurement bias of cognitive impairment being positively correlated with that of delirium, PU staging, and incontinence.

Table 4

The Impact of Facility Characteristics on Measurement Bias of 8 MDS Scales[a]

	Facility Characteristics							
Scale Differences	Run for Profit	Hospital Based	Has an Alzheimer's Unit	Number of Deficiencies > 50th Percentile in State	Part of a Chain	Total Registered Nurse Hours Per Patient Day > Nation's 75th Percentile in 2001	Total Admin Nurse Full Time Employee > Nation's 75th Percentile in 2001	Residents/ On-Staff Nurse Aid FTE >Nation's 75th Percentile in 2001
Activities of daily diving	0.086	0.662[b]	0.347	–0.293	–0.093	0.093	0.158	–0.205
Cognitive Performance Scale	–0.033	–0.109	–0.006	–0.150[b]	0.004	0.037	0.004	–0.032
Delirium	0.038	0.215[b]	0.301[b]	–0.192[b]	–0.025	–0.015	0.031	0.053
Bowel incontinence	0.009	0.052	–0.043	0.001	0.004	0.016	0.010	–0.052
Bladder incontinence	–0.052	0.006	–0.021	–0.010	0.050	–0.024	0.014	–0.034
Pressure ulcer staging	0.040	–0.055	0.027	0.003	0.039	0.020	0.022	0.000
Pain	–0.045	0.049	–0.017	–0.008	–0.051	0.065	0.084	0.037
Restraints	0.040	0.070	0.065	–0.050	–0.005	–0.005	–0.004	0.030

[a]Results are from the full multilevel multivariate models with resident characteristics, facility characteristics, and facility locations. Eight sets of regression coefficients (rows) were estimated for the 8 outcomes simultaneously. The interpretation of the regression coefficients was that compared to nursing homes in the referent group, the nursing homes in the comparison group on average overcoded (if the regression coefficient was positive) or undercoded (if the regression coefficient was negative) the health domain by x unit. MDS = Minimum Data Set.

[b]The regression coefficient was significant at $P < .01$.

Table 5

The Impact of Facility Location on Measurement Bias of 8 MDS Scales[a]

	Location of NHs[b]				
MDS Scales	CA	IL	MO	OH	PA
Activities of daily living	1.001[c]	–0.739[c]	0.018	1.909[c]	0.275
Cognitive Performance Scale	0.227[c]	0.074	0.124	0.575[c]	0.054
Delirium	0.120	–0.021	–0.580[c]	–0.061	–0.259[c]
Bowel incontinence	0.049	–0.032	–0.223[c]	–0.046	0.073
Bladder incontinence	0.137[c]	0.012	–0.136[c]	–0.010	0.007
Pressure ulcer staging	0.130[c]	0.093[c]	–0.028	0.033	0.050
Pain	0.401[c]	–0.020	–0.275[c]	0.509[c]	0.022
Restraints	–0.027	–0.066	–0.077	0.065	–0.176[c]

[a]Results are from the full multilevel multivariate models with resident characteristics, facility characteristics, and facility locations. Eight sets of regression coefficients (rows) were estimated for the 8 outcomes simultaneously. The interpretation of the regression coefficients was that compared to nursing homes in the referent group, the nursing homes in the comparison group on average overcoded (if the regression coefficient was positive) or undercoded (if the regression coefficient was negative) the health domain by x unit. MDS = Minimum Data Set; NH = nursing home; CA = California; IL = Illinois; MO = Missouri; OH = Ohio; PA = Pennsylvania.

[b]Nursing homes in Tennessee are the referent group.

[c]The regression coefficient was significant at $P < .01$.

After inclusion of resident characteristics in the model, we observed little or no reduction in the resident or facility level variances or covariances. This is consistent with the finding that few resident characteristics included in the model were statistically significant in explaining the variation in the measurement difference between the raters on all 8 scales. With the inclusion of facility characteristics we observed a reduction by 4 to 20 percentage points in the

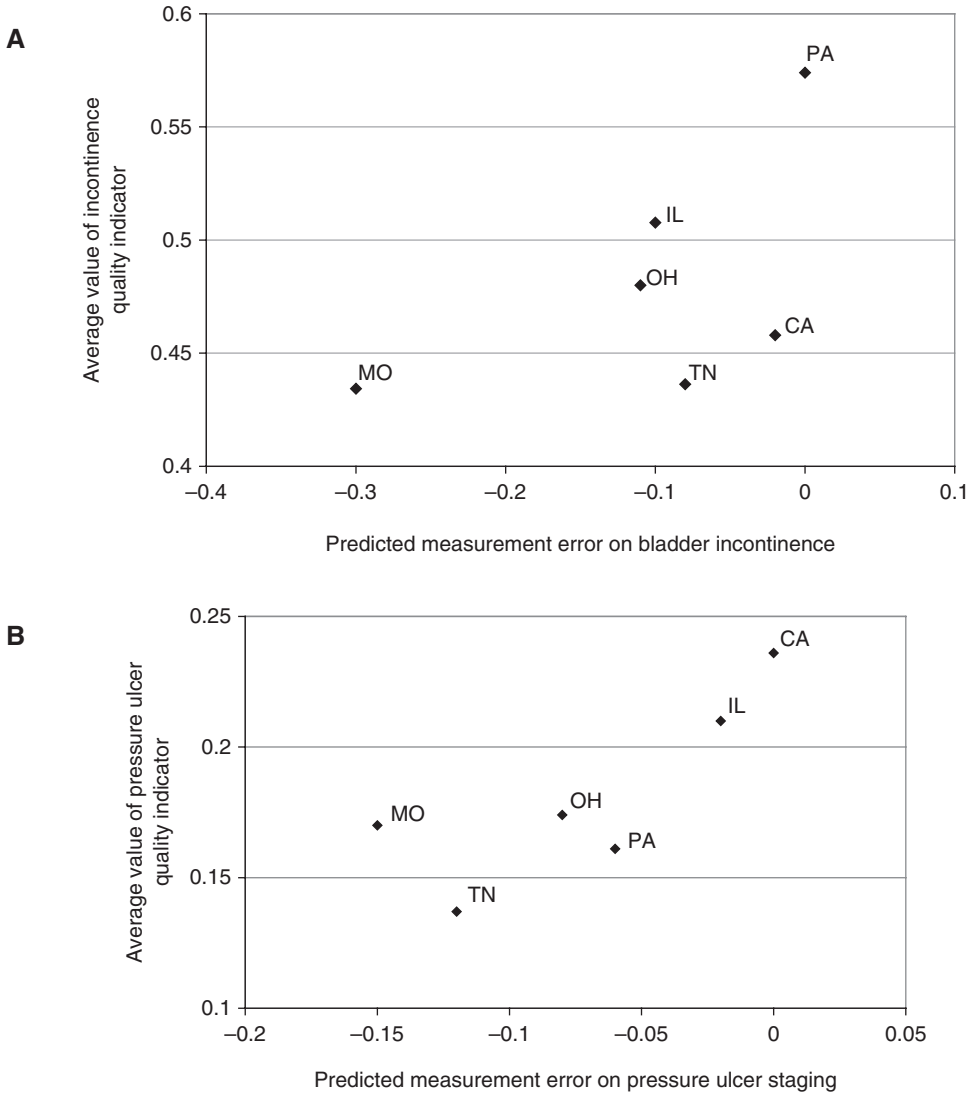

Figure 2. The association between state averages of quality indicators and the predicted values of measurement errors of MDS items that composed the quality indicators. (A) The association between state averages of incontinence quality indicator and the measurement errors on bladder incontinence. (B) The association between state averages of pressure ulcer quality indicators and the measurement errors on pressure ulcer staging. A lower value of a quality measure suggests the nursing home had better quality of care on the measured domain. MDS = Minimum Data Set; CA = California; IL = Illinois; MO = Missouri; OH = Ohio; PA = Pennsylvania; TN = Tennessee.

facility-level variances (Figure 1). Finally, we added the state in which facilities were located to the model and observed additional reduction by 13 to 34 percentage points in the facility-level variances.

From the full model, we found that nursing home staff tended to underrate levels of pain for residents 85 years of age or older, and overrate impairment of cognitive function in residents with dementia. Compared to nursing home staff in freestanding facilities, nursing home staff in hospital-based facilities were more likely to overdocument limitations in ADL and the level of delirium (Table 4). Staff, in facilities with more deficiencies cited during annual surveys than the state median, tended to underdocument the

impairment of residents' cognitive function or their level of delirium. No resident or facility characteristics, except for facility location, were simultaneously associated with 3 or more outcomes (Tables 4 and 5). Compared to facilities in Tennessee, facilities in California overrated ADL limitations, cognitive impairment, bladder incontinence, PU staging, and pain to a larger extent. Facilities in Missouri underrated delirium, bowel and bladder incontinence, and pain.

Figure 2 shows a positive association between the state averages of QIs and the measurement errors of MDS items that contribute to the calculation of the QIs. Compared to the participating facilities in California and Illinois, the participating facilities from Missouri, Tennessee, Ohio, and Pennsylvania underrated the stage of residents' PUs and had a lower prevalence of PUs. Similarly, facilities in Missouri underrated the severity of bowel incontinence and had the lowest average of percent of low-risk residents who lose control of bowel or bladder. This result suggests that undercoding of individual MDS items biased the QIs downward.

DISCUSSION

We found that the quality of facility coding for 8 MDS scales varied substantially across nursing homes. Variation among facilities in measurement errors can be partially explained by characteristics of the facility, but a substantial amount of this variation is due to state differences. This implies that there are state-level factors (ie, state policies, practices) that have an impact on measurement error. We also found that a generalized effect of nursing home state location tended to be consistent across MDS measures.

One or more factors may contribute to the variation of data quality across states. Each state has a branch of the National Association of Assessment Nurses that provides training and technical assistance for MDS data collection. The individuals who coordinate training programs may have their own interpretations of the assessment protocol. Thus, the state or regional interpretations may deviate from one another in the absence of a comprehensive centralized interpretation. Such deviation may have contributed to the variation in data quality across states.

State policies that regulate nursing home care may also affect data quality. Facilities in states with more stringent survey processes and policies may underdocument adverse events to avoid sanctions.[21] Nursing homes in states with a case mix-based reimbursement system may have a higher incentive to overreport conditions that lead to more reimbursement.[22] Because we have reliability data from only 6 states, we are unable to test these theories. If state policies cause the variation in data quality, then we should be concerned about how data inaccuracy affects surveillance and reimbursement, and what measures can be taken to improve data quality.

We also found that the direction and magnitude of measurement bias may undermine the validity of the publicly reported QIs. Nursing homes in states that tend to have overrated residents' problems had higher QIs than those that underrated them (Figure 2A and 2B). Moreover, because the state effect on measurement bias tended to be consistent across measures, measurement bias poses a larger threat to composite measures or case mix index that are derived from more MDS items. Instead of canceling each other, these measurement errors in the same direction would tend to accumulate in composite measures. If such an association exists, there are important policy implications. It is difficult to draw conclusions based on the reliability data collected from 6 states. Further studies on larger samples are needed.

Quality indicators of nursing home care are publicly released to assist consumers in selecting providers. If facilities that undercoded residents' problems had lower QIs (suggesting good quality of care), consumers would be misled by the published QIs. This impacts consumers who pick facilities across state lines more than it impacts those who choose facilities within the same market area. Cross-state selections often happen when consumers must decide whether to live in a facility close to their own homes or close to their children's homes.

While building the multilevel models, we tested the following resident and facility characteristics that we felt might be associated with the quality of MDS data: residents' primary language, marital status, having cancer or not, facility quality of care performance, number of beds per facility,

percentage of residents on Medicare, whether the facility had a certified nurse aid training program, and whether the facility was in an urban or rural area. The inclusion of these variables did not change the model performance or our conclusions. Consequently, we did not keep them in the full model. Our results were consistent with previous findings. Mor and colleagues studied the association between facility characteristics and reliability of MDS items. No systematic facility-level association was identified.[13]

We suspected that skilled nurse staffing levels might be associated with discrepancies, but our results did not show such an association. We further tested different ways to define the facility staffing level (either absolute values or indicators greater than state median). This, too, was not born out in the analysis. It may be the case that known problems with the staffing data quality (annual OSCAR facility survey reported at the time of facility inspection) may have reduced the sensitivity of our test.[23] Another alternative explanation is that the staffing turnover rate, instead of staffing level, is related to the quality of MDS data. High turnover rates interfere with the continuity of care and redirect resources available for providing care to other activities such as staff training.[24,25] However, turnover rate is not routinely collected and was not available for this study.

We treated the data collected by research nurses as the gold standard. However, research nurses could make mistakes just as facility staff could. Compared to nursing home staff, research nurses had less time to interact with residents, and so might have underdocumented some incidents and symptoms. Study nurses and nursing home staff relied on the medical chart to fill out the MDS forms. Different interpretations of the charts between the study nurses and nursing home staff could lead to a disagreement about the MDS coding. On the other hand, study nurses were uniformly trained and their MDS data from multiple nursing homes were highly reliable. In contrast, the assessments done by facility staff were not specifically collected for the QI validation study but were part of the routine assessments conducted by nursing homes to fulfill the US government's mandate. Therefore, these can be assumed to reflect the real quality of MDS data available to the US government. Thus, we attribute the coding differences between nursing home staff and study nurses to differential adherences to RAI protocol and random measurement errors.

While estimating the reliability of MDS items, we assumed that resident status remained the same and attributed coding differences between nursing home staff and study nurses to coding errors. In fact, some residents may have improved or deteriorated in the measured health domains during the interval between the paired assessment (average interval, 23 days; standard deviation, 19 days), and consequently we would underestimate the reliability of MDS coding. On the other hand, we do not expect that our results are unduly biased. Most MDS items document resident status during the 7 days prior to the assessment rather than the status observed at 1 time point and thus are less variable. We expect that MDS coding would remain the same over a 4-week period for a majority of residents, especially for the long-stay residents who lived in nursing homes for years.

SUMMARY

We found substantial variations in MDS measurements among nursing homes that cannot be explained by the case mix of nursing home residents or by the characteristics of nursing homes that we had available in our data set. This variation could be explained partially by the state in which the nursing home is located. State policy on surveillance or reimbursement may contribute to the variation. State associations and training organizations also may introduce a systematic shift in the interpretation of assessment criteria, thus contributing to the variation. Unless we adjust for, or eliminate, the substantial variation in measurement across states, comparison of quality of care and payment based on performance will be adversely affected. Given the potential such systematic measurement bias has for undermining the validity of public reports of quality as well as of pay-for-performance schemes, it is essential that the US federal government undertake a comprehensive review of the uniformity of MDS assessment practices and institute ongoing centralized data monitoring and audit programs.

Appendix

Scales Derived From Items in Minimum Data Set

Scales	Score Range	MDS Items That Composite the Scale	Average Correlation Between MDS Items
Activities of daily living (ADL)	0–28	**g1aa.** How a resident moves to and from lying position, turns side to side, and positions body while in bed **g1ba.** How a resident moves between surfaces—to and from bed, chair, wheelchair, and standing position (exclude to and from the bath/toilet) **g1ea.** How a resident moves between locations in his/her room and adjacent corridor on same floor; if in wheelchair, self-sufficiency once in chair **g1ga.** How a resident puts on, fastens, and takes off all items of street clothing, including donning and removing prosthesis **g1ha.** How a resident eats and drinks (regardless of skill); includes intake of nourishment by other means (eg, tube feeding, total parenteral nutrition) **g1ia.** How a resident uses the toilet room (or commode, bedpan, urinal); transfer on and off toilet, cleanses, changes pad, manages ostomy or catheter, adjusts clothes **g1ja.** How a resident maintains personal hygiene, including combing hair, brushing teeth, shaving, applying makeup, washing and drying face, hands, and perineum (exclude baths and showers)	0.61
Cognitive Performance Scale (CPS)	0–6	**b1.** Comatose (persistent vegetative state/no discernible consciousness) **g1aa.** How a resident moves to and from lying position, turns side to side, and positions body while in bed **g1ba.** How a resident moves between surfaces—to and from bed, chair, wheelchair, standing position (exclude to and from the bath/toilet) **g1ha.** How a resident eats and drinks (regardless of skill); includes intake of nourishment by other means (eg, tube feeding, total parenteral nutrition) **g1ia.** How a resident uses the toilet room (or commode, bedpan, urinal); transfer on and off toilet, cleanses, changes pad, manages ostomy or catheter, adjusts clothes **b2a.** Memory: short-term memory alright—seems/appears to recall after 5 minutes **b4.** Cognitive skills for daily decision making (made decisions regarding tasks of daily life) **c4.** Making self understood (expressing information content however able)	0.39
Pain	0–6	**j2a.** Frequency with which resident complains or shows evidence of pain **j2b.** Intensity of pain	0.89
Restraints	0–6	**p4c.** Devices and restraints trunk restraint **p4d.** Devices and restraints limb restraint **p4e.** Devices and restraints chair prevents rising	0.22
Bowel incontinence	0–4	**h1a.** Control of bowel movement, with appliance or bowel continence programs, if employed	
Bladder incontinence	0–4	**h1b.** Control of urinary bladder function (if dribbles, volume insufficient to soak through underpants), with appliances (eg, foley) or continence programs, if employed	
Pressure ulcer staging	0–4	**m2a.** The highest stage of pressure ulcers in the last 7 days	
Delirium	0–12	**b5a.** Easily distracted (eg, difficulty paying attention, gets sidetracked) **b5b.** Periods of altered perception or awareness of surroundings (eg, moves lips or talks to someone not present, believes he/she is somewhere else, confuses night and day) **b5c.** Episodes of disorganized speech (eg, speech is incoherent, nonsensical, irrelevant, or rambling from subject to subject; loses train of thought) **b5d.** Periods of restlessness (eg, fidgeting or picking at skin, clothing, napkins; frequent position changes; repetitive physical movements or calling out) **b5e.** Periods of lethargy (eg, sluggishness, staring into space, difficult to arouse, little body movement) **b5f.** Mental function varies over the course of the day (eg, sometimes better, sometimes worse; behaviors sometimes present, sometimes not)	0.37

For all the scales, a higher value suggests more severe impairment in the measured health domain. MDS = Minimum Data Set.

REFERENCES

1. Achterberg WP, Holtkamp CC, Kerkstra A, Pot AM, Ooms ME, Ribbe MW. Improvements in the quality of coordination of nursing care following implementation of the resident assessment instrument in Dutch nursing homes. *J Adv Nurs.* 2001;35:268-275.
2. Brunton B, Rook M. Implementation of the resident assessment instrument: a Canadian experience. *Healthc Manage Forum.* 1999;12:49-53.
3. Hawes C, Morris JN, Phillips CD, Fries BE, Murphy K, Mor V. Development of the nursing home resident assessment instrument in the USA. *Age Ageing.* 1997;26:19-25.
4. Mor V. A comprehensive clinical assessment tool to inform policy and practice: applications of the Minimum Data Set. *Med Care.* 2004;42:III50-III59.
5. Harris Y, Clauser SB. Achieving improvement through nursing home quality measurement. *Health Care Financ Rev.* 2002;23:5-18.
6. Grabowski DC. The economic implications of case-mix Medicaid reimbursement for nursing home care. *Inquiry.* 2002;39:258-278.
7. Reilly KE, Mueller C, Zimmerman DR. The Centers for Medicare and Medicaid Services' nursing home case-mix and quality demonstration: a descriptive overview. *J Aging Soc Policy.* 2007;19:61-76.
8. Lawton MP, Casten R, Parmelee PA, Van Haitsma K, Corn J, Kleban MH. Psychometric characteristics of the Minimum Data Set II: validity. *J Am Geriatr Soc.* 1998;46:736-744.
9. Casten R, Lawton MP, Parmelee PA, Kleban MH. Psychometric characteristics of the Minimum Data Set I: confirmatory factor analysis. *J Am Geriatr Soc.* 1998;46:726-735.
10. Morris JN, Moore T, Jones R, et al. Validation of long-term and postacute care quality indicators. CMS contract No: 500-95-0062/T.O. #4. Available at: http://www.interrai.org/applications/qireport2.pdf. Accessed February 28, 2008.
11. 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;35:172-178.
12. Bates-Jensen BM, Cadogan M, Osterweil D, et al. The Minimum Data Set pressure ulcer indicator: does it reflect differences in care processes related to pressure ulcer prevention and treatment in nursing homes? *J Am Geriatr Soc.* 2003;51:1203-1212.
13. Mor V, Angelelli J, Jones R, Roy J, Moore T, Morris JN. Interrater reliability of nursing home quality indicators in the US. *BMC Health Serv Res.* 2003;3:20.
14. Schnelle JF, Wood S, Schnelle ER, Simmons SF. Measurement sensitivity and the Minimum Data Set depression quality indicator. *Gerontologist.* 2001;41:401-405.
15. Simmons SF, Cadogan MP, Cabrera GR, et al. The Minimum Data Set depression quality indicator: does it reflect differences in care processes? *Gerontologist.* 2004;44:554-564.
16. Wu N, Miller SC, Lapane K, Gozalo P. The problem of assessment bias when measuring the hospice effect on nursing home residents' pain. *J Pain Symptom Manage.* 2003;26:998-1009.
17. Wu N, Miller SC, Lapane K, Roy J, Mor V. The quality of the quality indicator of pain derived from the Minimum Data Set. *Health Serv Res.* 2005;40:1197-1216.
18. Lum TY, Lin WC, Kane RL. Use of proxy respondents and accuracy of Minimum Data Set assessments of activities of daily living. *J Gerontol A Biol Sci Med Sci.* 2005;60:654-659.
19. Fries BE, Simon SE, Morris JN, Flodstrom C, Bookstein FL. Pain in U.S. nursing homes: validating a pain scale for the Minimum Data Set. *Gerontologist.* 2001;41:173-179.
20. Sangl J, Saliba D, Gifford DR, Hittle DF. Challenges in measuring nursing home and home health quality: lessons from the First National Healthcare Quality Report. *Med Care.* 2005;43:I24-I32.
21. Harrington C, Carrillo H. The regulation and enforcement of federal nursing home standards, 1991-1997. *Med Care Res Rev.* 1999;56:471-494.
22. Chen LW, Shea DG. Does prospective payment really contain nursing home costs? *Health Serv Res.* 2002;37:251-271.
23. Feng Z, Katz PR, Intrator O, Karuza J, Mor V. Physician and nurse staffing in nursing homes: the role and limitations of the Online Survey Certification and Reporting (OSCAR) system. *J Am Med Dir Assoc.* 2005;6:27-33.
24. Donoghue C, Castle NG. Organizational and environmental effects on voluntary and involuntary turnover. *Health Care Manage Rev.* 2007;32:360-369.
25. Castle NG, Engberg J, Men A. Nursing home staff turnover: impact on nursing home compare quality measures. *Gerontologist.* 2007;47:650-661.