Impact of Medicaid Reimbursement on Mental Health Quality Indicators

Nicole M. Bellows and Helen A. Halpin

Objective. To examine the relationship between the use of the Minimum Data Set (MDS) for determining Medicaid reimbursement to nursing facilities and the MDS Quality Indicators examining nursing facility residents' mental health.

Data Sources. The 2004 National MDS facility Quality Indicator reports served as the dependent variables. Explanatory variables were based on the 2004 Online Survey Certification and Reporting system (OSCAR) and an examination of existing reports, a review of the State Medicaid Plans, and State Medicaid personnel.

Study Design. Multilevel regression models were used to account for the hierarchical structure of the data.

Data Collection. MDS and OSCAR data were linked by facility identifiers and subsequently linked with state-level variables.

Principal Findings. The use of the MDS for determining Medicaid reimbursement was associated with higher (poorer) quality indicator values for all four mental health quality indicators examined. This effect was not found in four comparison quality indicators.

Conclusions. The findings indicate that documentation of mental health symptoms may be influenced by economic incentives. Policy makers should be cautioned from using these measures as the basis for decision making, such as with pay-for-performance initiatives.

Key Words. Medicaid, mental health, quality of care, long-term care

---

The Minimum Data Set (MDS) is an important measurement tool used by the Centers for Medicare and Medicaid Services (CMS) and state health regulators. MDS data are collected on all nursing facility residents in Medicare and Medicaid certified facilities and are used for two main purposes: (1) to determine the appropriate daily case-mix nursing facility reimbursement rate for Medicare and several state Medicaid programs via the Resource Utilization Group (RUG-III) system, and (2) to create the MDS Quality Indicators (QIs).

The QIs were developed at the University of Wisconsin-Madison's Center for Health Systems Research and Analysis (CHSRA) and created with

the intent to assess quality and foster quality improvement in nursing facilities (Zimmerman et al. 1995). The QIs are based on specific person-level MDS data aggregated to reflect the percentage of nursing facility residents with specific undesirable health conditions (e.g., depression symptoms) or potentially poor service indicators (e.g., the use of restraints). Because higher QI values represent poorer quality, the MDS QIs are referred to as “poor quality indicators” or “poor QI values” from this point forward.

A wide variety of information on nursing facility residents is captured by the MDS, including aspects of the residents’ mental health and well-being such as cognitive status and the presence of depressive symptoms and behavioral symptoms affecting others. These measures are used in the RUG-III case-mix reimbursement system, where residents who are identified as having depression symptoms, behavioral symptoms affecting others, and cognitive impairment may have higher daily reimbursement rates assigned to them compared with similar residents who are not identified as having these symptoms.

These same MDS data are also used in generating four of the poor quality indicators: (1) prevalence of behavioral symptoms affecting others, (2) prevalence of symptoms of depression, (3) prevalence of symptoms of depression with no antidepressant therapy, and (4) incidence of cognitive impairment. Supplementary material Appendix A1 describes how these measures are calculated and the MDS components used to generate these measures.

CMS has provided data on the QIs to the general public where state-level measures of the QIs are available on the CMS website and similar facility measures, based on the same MDS data as the QIs, are available on the Medicare Nursing Home Compare website. Additionally, The CMS website currently states that the agency is considering using the QIs or the related quality measures in a “pay-for-performance” demonstration initiative in nursing facilities (CMS 2006). Pay-for-performance initiatives aim to improve the quality of care through financial incentives, based on quantifiable measures (Rowe 2006), however, some have challenged whether pay-for-performance strategies with fixed performance targets can effectively improve quality (Rosenthal et al. 2005).

Address correspondence to Nicole M. Bellows, Ph.D., Assistant Research, Center for Health and Public Policy Studies, University of California, Berkeley, 140 Warren Hall #7360, Berkeley, CA 94720-7360. Helen A. Halpin, Ph.D., Professor of Health Policy, Center for Health and Public Policy Studies, University of California, Berkeley, Berkeley, CA.

While some researchers have found that most of the nursing facility QIs have sufficient validity and reliability to be used in research (Karon, Sainfort, and Zimmerman 1999; Morris et al. 2003), other researchers have challenged the validity and reliability of the QIs and have suggested that while the QIs have been used for regulatory purposes, they are still very much a work in progress (Arling et al. 2005; Mor 2005; Sangl et al. 2005).

A further challenge to these measures is that they could be artificially inflated in states with MDS-based case-mix reimbursement systems in response to financial incentives to document nursing home residents with higher acuity, where symptoms of behaviors affecting others, depression symptoms, and poor cognitive functioning are factors in determining resident acuity. Researchers have found that one of the major responses to states adopting a case-mix reimbursement system is that access to nursing facilities increased for higher acuity residents (Arling and Daneman 2002; Grabowski 2002). Some researchers have suggested, however, that the increases in patient acuity after case-mix implementation may have been due to documentation changes rather than actual changes in case mix (Arling and Daneman 2002; Weissert and Muslinger 1992).

The documentation changes can be the result of poor past documentation being corrected or a more aggressive profit-oriented documentation strategy of “gaming” (Lu 1999; Courty and Marschke 2003). Health economists examined gaming when Medicare changed to a prospective payment system for hospital services. Researchers found that at least part of the increase in case mix among hospitals after the Diagnosis Related Groups (DRG) system was introduced was attributed to differences in documentation or “DRG creep” (Steinwald and Dummit 1989; Carter, Newhouse, and Relles 1990; Hsia et al. 1992; Silverman and Skinner 2004).

The experience of “DRG creep” is particularly relevant to the use of the MDS for nursing facility policy because MDS data are the basis for case-mix reimbursement in the RUG-III systems and they are also the source for the poor QIs. One report by the Office of Inspector General (2001) found that both “upcoding” and “downcoding” were occurring in the MDS assessments of Medicare nursing facility residents and suggested that these findings reflect confusion over the assessment process, rather than systematic bias. Aside from this report, however, there is a lack of published research on upcoding or downcoding of the MDS.

One area that allows an investigation of documentation incentives in nursing facilities is the differences in state Medicaid programs and their use of the MDS for reimbursement. Because Medicaid is the largest payer of nursing

facility care (U.S. GAO 2002a), nursing facilities located in states with a MDS-based Medicaid reimbursement system, referred to in this article as “MDSM states,” have a greater incentive to document these symptoms because documentation can result in higher payment amounts. Therefore, this paper examines whether facilities located in states that use the MDS for state Medicaid reimbursement have higher poor QI values compared with facilities in other states.

DESIGN AND METHODS

Study Sample

The study sample included facilities within the 48 contiguous United States with at least 20 residents. Medicare-only certified facilities were not included in the sample because they would not be directly affected by incentives in the Medicaid reimbursement systems. Also eliminated from the data set were another 100 facilities that reported more residents than certified beds, therefore suggesting erroneous data (Konetzka et al. 2004). After these restrictions, approximately 13,000 nursing facilities had complete OSCAR and MDS QI data. The sample size for each of the four dependent variables analyzed ranged from 12,847 to 12,997.

Dependent Variables

The four mental health poor QIs were selected for analysis because the underlying MDS data are also used in the RUG-III reimbursement systems. In addition to the mental health poor QIs, four comparison poor QIs were examined where the underlying MDS data for these measures are not used in the RUG-III reimbursement system and therefore one would not expect an upcoding effect, in contrast to the mental health poor QIs. The four comparison QIs are: prevalence of urinary tract infections, prevalence of indwelling catheters, prevalence of use of nine or more different medications, and prevalence of any antianxiety or hypnotic use. Supplementary material Appendix A1 describes these measures in more detail.

The dependent variables were obtained through the CMS National MDS facility QI numerator/denominator data reports for 2004. CMS generated monthly numerator (number of residents with specified symptoms) and denominator (total residents) data. The percentages of facility residents with each poor QI were calculated for each month and yearly averages were calculated and used as the dependent variables to control for monthly instability

in the QIs (Mor et al. 2003). Because the dependent variables had substantially right-skewed distributions, the dependent variables were transformed by taking natural log of the percentage values plus one, where the plus one was used in order to retain observations with zero percentage values. The exception was for the dependent variable that captured the use of nine or more different medications, which had a normal distribution.

Explanatory Variables

The primary explanatory variable was a state-level binary variable distinguishing whether or not the state had an MDS-based Medicaid reimbursement system in 2004 (MDSM). The use of the MDS data for both reimbursement and quality-monitoring purposes was hypothesized to be associated with state-level differences in the four poor QI values. An alternative explanation is that it is not the explicit use of the MDS that may result in state-level QI differences, but rather that the use of a case-mix reimbursement methodology results in state-level differences. To control this possibility, states that did not have MDS-based Medicaid reimbursement systems were separated into two groups, those that used a case-mix system that was not based on the MDS but on a different data source (Other Case Mix) and those that did not use any case mix for Medicaid reimbursement (No Case Mix).

The MDSM, Other Case Mix, and No Case Mix variables were obtained through existing reports that detail the structure of Medicaid reimbursement systems (Harrington et al. 2000; U.S. GAO 2002b) and a review of the State Medicaid Plans and Plan Amendments located on the CMS website. State Medicaid personnel confirmed the coding of these variables. Table 1 details the classification of each state among the three Medicaid reimbursement system categories. Twenty-five states were classified as MDSM, 10 states were classified as Other Case Mix, and 13 states were classified as No Case Mix.

Another state-level explanatory variable examined was the average state daily nursing facility Medicaid rate. Grabowski et al. (2004) reported the average state Medicaid rates for 2002, which were inflated to 2004 dollars using the skilled nursing facility input price index (CMS 2007). The average daily state Medicaid rate for 2004 ranged from \$89 to \$185 with a mean of approximately \$127 and a standard deviation of \$23.

In addition to the state-level variables, seven facility-level explanatory variables were also examined including: total licensed nurse staffing hours per patient day, certified nursing assistant (CNA) staffing hours per patient day,

Table 1: Classification of 48 States in the Medicaid Reimbursement System in 2004

<i>MDS-Based Case Mix (MDSM)</i>		<i>Other Case Mix (Other CM)</i>	<i>No Case Mix (No CM)</i>
Colorado	Nebraska	Arizona	Alabama
Georgia	Nevada	Delaware	Arkansas
Idaho	New Hampshire	Illinois	California
Indiana	North Carolina	Maryland	Connecticut
Iowa	North Dakota	Massachusetts	Florida
Kansas	Ohio	New Jersey	Michigan
Kentucky	Pennsylvania	New York	Missouri
Louisiana	South Dakota	Oregon	New Mexico
Maine	Utah	South Carolina	Oklahoma
Minnesota	Vermont	Texas	Rhode Island
Mississippi	Virginia		Tennessee
Montana	Washington		Wisconsin
	West Virginia		Wyoming
25 states		10 states	13 states

facility ownership (for-profit versus nonprofit or government), facility size, occupancy rate, percent of facility residents reimbursed through Medicaid, and metropolitan versus nonmetropolitan location. The Online Survey, Certification and Reporting System (OSCAR) data served as the data source for these facility-level variables. Table 2 details the summary statistics for each of the facility-level explanatory variables.

Analysis Approach

This study examines cross-sectional data with the hopes of pursuing a longitudinal analysis in the future. Multilevel or “mixed” regression models were used to account for the hierarchical structure of the data (facilities nested within states) and produce correct standard errors (Rabe-Hesketh and Skrondal 2005). All statistical analyses were conducted using *STATA 9.1*. Random-intercept multilevel models were estimated first, followed by random-coefficient multilevel models that allowed the effects of the for-profit and the high percent Medicaid variables to vary randomly over states (individually and with both random coefficients). Likelihood ratio tests were used to determine the models that best fit the data.

Because r^2 and adjusted r^2 values are not generated when using multilevel models in *STATA*, an “estimated r^2 ” was calculated for each of the final models. To estimate this measure, the variances from the final models with the

Table 2:   Summary Statistics for Facility-Level Explanatory Variables

<i>Variable</i>	<i>Distribution</i>	<i>Summary Statistics</i>
Licensed nurse staffing per patient per day	Continuous, normal	Mean = 1.21 Standard deviation = 0.34
CNAs per patient per day	Continuous, normal	Mean = 2.28 Standard deviation = 0.57
Percent Medicaid residents	Categorical (low, average, high)	Mean = 66.10 Standard deviation = 17.80 Low: <60% = 30% Average: 60–80% = 46% High: >80% = 24%
For-profit ownership	Binary (For-profit = 1)	72% for-profit 28% nonprofit or government
Nonmetro location	Binary (Nonmetro = 1)	31% nonmetro 69% metro
Size of facility (certified beds)	Categorical (small, medium, large)	Mean = 113 Standard deviation = 61 Low: <75 beds = 28% Medium: 75–149 beds = 52% High: ≥ 150 beds = 20%
Occupancy rate	Categorical (low, average, high)	Mean = 86 Standard deviation = 13 Low: <80% = 24% Average: 80–94% = 47% High: ≥ 95% = 29%

included covariates were subtracted from the variances of the random intercept models without covariates. The differences were then divided by the variances of the random intercept models without the covariates.

RESULTS

Table 3 details the models that best fit the data for each of the eight dependent variables analyzed. The extent to which the models explained variation in the QIs differed by QI with the estimated r^2 values ranging from a low of 0.057 to a high of 0.164.

The MDSM variable had a positive statistically significant association with all four mental health poor QIs, indicating that facilities located in MDSM states had higher poor QI values compared with facilities located in non-MDSM states. In contrast, the MDSM variable was not statistically significantly associated with any of the comparison poor QI measures, where no

Table 3: Multi-Level Regression Models for Mental Health Poor Quality Indicators

<i>Variable</i>	<i>Model 1: Behavioral Symptoms (N = 12,997)</i>	<i>Model 2: Depression Symptoms (N = 12,905)</i>	<i>Model 3: Depression Symptoms w/o Antidepressants (N = 12,997)</i>	<i>Model 4: Cognitive Impairment (N = 12,847)</i>
Constant	2.715***	2.009***	1.436***	2.238***
<i>State variables</i>				
MDSM	0.277***	0.492***	0.488***	0.093*
Other case mix	0.068	-0.023	-0.017	-0.015
Average state Medicaid nursing facility rate	-0.000	0.000	-0.001	-0.001
<i>Facility variables</i>				
Licensed nurse staff per patient per day	-0.079***	0.063**	0.038	0.107***
CNA staff per patient per day	0.021*	0.003	0.003	0.002
For-profit	0.039*	-0.025	-0.023	0.058***
Nonmetropolitan location	0.063***	0.068***	0.106***	-0.043**
Low percent Medicaid residents	-0.122***	-0.023	-0.022	0.106***
High percent Medicaid residents	0.147***	-0.013	-0.012	-0.167***
Small size	0.035**	0.072***	0.037*	-0.111***
Large size	0.044**	-0.012	0.007	0.019
Low occupancy	0.088***	0.060***	0.054**	-0.049**
High occupancy	0.020	0.034*	0.026	0.018
<i>Random coefficients</i>				
Random coefficients included in model	High percent Medicaid residents, for-profit	High percent Medicaid residents, for-profit	None	High percent Medicaid residents
<i>Measure of fit</i>				
Estimated R^2	0.112	0.164	0.133	0.057

* $p < .05$,

*** $p < .01$,

**** $p < .001$.

Omitted categories include: "No Case Mix," not-for-profit/government ownership, metropolitan location, average percent Medicaid residents, medium size, average occupancy.

financial incentives existed to document the underlying MDS components for higher case-mix reimbursement (Table 4).

Because the mental health poor QIs were log transformed and the explanatory variables were not, the MDSM variable results in a $100 \times (\beta \text{ coefficient})$ percent change for the mental health poor QIs, holding all other variables constant. As such, the use of the MDS for state Medicaid reimbursement resulted in a 9 percent increase in the incidence cognitive impairment QI, a 28 percent increase in the prevalence of behavioral symptoms affecting others, a 49 percent increase in the prevalence of depression symptoms without antidepressants, and a 49 percent increase in the prevalence depression symptoms.

Being located in a state that used a case-mix Medicaid reimbursement system based on a different data source (Other Case Mix) was not statistically significantly associated with any of the poor QIs. Additionally, the average Medicaid reimbursement rate was only statistically significantly associated with the prevalence of antianxiety/hypnotic use QI.

The direction of the association and level of statistical significance of the nine facility-level variables differed across the eight QIs. For example, the small size variable had a statistically significant positively association with three of the QIs and a statistically significant negative association with five of the QIs. Similar patterns of statistically significant relationships in countering directions were evident for most of the facility-level explanatory variables.

DISCUSSION

As mentioned previously, the mental health poor QIs are particularly interesting measures to examine because the underlying MDS data are used for both reimbursement (via RUG-III reimbursement systems) and quality monitoring (via the poor QIs). Additionally, the mental health QIs are important to investigate due to questions on their appropriateness for use as quality measures. Researchers who have investigated the prevalence of mental health conditions in nursing facilities have found them to be underreported in nursing facility records, particularly for measures of depression (Jones, Marcan-tonio, and Rabinowitz 2003; Simmons et al. 2004). In fact, Schnelle et al. (2001) conclude that the prevalence of depression symptoms QI may be a better indicator of depression recognition and documentation than of quality of care. As a result, it may be inappropriate to consider these measures as quality measures because it can potentially penalize facilities that document

Table 4: Multi-Level Regression Models for Comparison Poor Quality Indicators

<i>Variable</i>	<i>Model 5: UTIs (N = 12,997)</i>	<i>Model 6: Indwelling Catheters (N = 12,997)</i>	<i>Model 7: Antianxiety/ Hypnotic Use (N = 12,991)</i>	<i>Model 8: 9+ Meds (N = 12,997)</i>
Constant	1.908***	1.843***	3.014***	54.777***
<i>State variables</i>				
MDSM				1.750
Other case mix	0.021	0.079	0.017	— 0.96
Average state Medicaid nursing facility rate	— 0.022	— 0.092	— 0.045	— 0.040
Average state Medicaid nursing facility rate	— 0.000	— 0.001	— 0.002*	
<i>Facility variables</i>				
Licensed nurse staff per patient per day	0.155***	0.196***	0.045***	5.288***
CNA staff per patient per day	0.046***	0.011	0.016	0.628**
For-profit	0.087***	0.145***	0.011	— 0.575
Nonmetropolitan location	0.009	0.018	0.099***	1.655***
Low percent Medicaid residents	0.051**	0.002	— 0.008	0.511*
High percent Medicaid residents	— 0.225***	— 0.177***	— 0.032	— 2.782***
Small size	— 0.061***	— 0.192***	— 0.033**	— 0.804**
Large size	— 0.011	0.047**	— 0.017	— 0.951***
Low occupancy	— 0.028*	0.028*	0.034**	— 0.885**
High occupancy	— 0.008	— 0.069***	0.007	0.808**
<i>Random coefficients</i>				
Random coefficients included in model	High percent Medicaid residents	High percent Medicaid residents, for-profit	High percent Medicaid residents, for-profit	High percent Medicaid residents, for-profit
<i>Measure of fit</i>				
Estimated R^2	0.110	0.092	0.062	0.124

* $p < .05$,

*** $p < .01$,

*** $p < .001$.

Omitted categories include: “No Case Mix,” not-for-profit/government ownership, metropolitan location, average percent Medicaid residents, medium size, average occupancy.

these conditions by profiling them as having poorer quality. This study contributes to this discussion by examining a state-level factor that might influence the documentation of poor mental health symptoms.

The primary finding of this research is that for each of the four mental health poor QIs, a “MDSM effect” was found where facilities located in states that used the MDS in their Medicaid reimbursement system had statistically significantly higher poor QI values compared with facilities located in non-MDSM states. It is likely that because facilities in MDSM states relied on MDS documentation for Medicaid reimbursement they had an increased incentive to document these poor mental health symptoms compared with non-MDSM states, resulting in different QI values. In contrast, no statistically significant MDSM effect was found for the four comparison QIs where there is not an economic incentive to document the health conditions underlying the comparison QIs.

These findings suggest the possibility that facilities in MDSM states are upcoding the mental health MDS data (making residents appear to have more depression symptoms, cognitive impairment, and having more behavioral symptoms than actuality) in order to capture more reimbursement. An alternative explanation is that facilities in non-MDSM states could be downcoding the data (not documenting these symptoms when present) due to a lack of an administrative prerogative to complete the MDS or more deliberately in order to keep their poor QI values low. Finally, it is possible that due to the financial incentives in MDSM states, there is better access to nursing facility care for residents with these mental health symptoms in MDSM states.

Regardless of the reason why there is this MDSM and non-MDSM difference, these findings have important policy implications. First, if the difference is due to miscoding practices (upcoding or downcoding) among nursing facilities, it threatens the integrity of the case-mix process, which aims to pay more for more resource dependent residents. Within MDSM states, facilities that accurately assess mental health symptoms may be penalized financially if their competitors are systematically upcoding the MDS data. Additionally, because the RUG-III system is also used for Medicare reimbursement, miscoding MDS data also threatens the integrity of the Medicare reimbursement process. Furthermore, miscoding the MDS data also threatens the validity of the poor QIs and the ability of regulators and consumers to compare facilities based on these measures.

It is necessary to acknowledge this study’s limitations. First, this study uses cross-sectional data to confirm a hypothesis regarding coding differences

of the MDS between MDSM and non-MDSM states. As such, the results can only indicate a correlation, not causation. Future research should examine longitudinal data that could test for temporal changes in quality indicators in states that switched to an MDS-based reimbursement system in recent years.

A second important limitation is a general concern cited in the literature regarding the use of administrative data for long-term care, which may contain problems with validity (Ryan, Stone, and Raynor 2004). There are specific limitations with the dependent variables that are used in this analysis. For example, many researchers have challenged the validity and reliability of the depression symptoms measure from the MDS and found it to have low correlations with competing depression scales (Frederickson, Tariol, and DeJonghe 1996; Schnelle et al. 2001; Snowden 2004). While these concerns are important, the main hypothesis of this study is that the Medicaid reimbursement system influences the identification and documentation of these conditions and this study does not make claims about the accuracy of the data for specific diagnoses. In fact, the purpose of this study is to identify whether reimbursement incentives influence identification of these symptoms in nursing facilities and therefore lend further challenges to the validity of the MDS-based measures reflected in the QIs related to mental health.

Another limitation is that it is possible that states' decisions to use MDS-based case-mix adjustments are not random, allowing for potential unobserved variable bias. In order to address this issue, the MDSM status of the states was examined for correlations with multiple other state-level variables regarding the size of the Medicaid program, the age and income levels of the states' population, and other state policy variables. The only state-level variable found to be statistically significantly correlated with MDSM status was the percent of the population that lived in nonmetropolitan areas, resulting in the inclusion of the nonmetropolitan facility-level variable in the final models. Other state-level variables not examined could also contribute to explaining variation in the dependent variables. Still, one advantage of the multilevel model is that it accounts for the possibility that not all relevant state-level variables were specified in the model because the multilevel models assume that facilities operating in the same state are correlated after controlling for covariates.

What is not clear from this research is whether the economic incentives embedded in the Medicaid system for MDSM states improves data accuracy. It is quite possible that the conflicting incentives for MDS documentation, when these measures are used for both reimbursement and quality monitoring, result in better data accuracy as opposed to having them used for only one

purpose. As such, CMS should commission an independent evaluation regarding the implications of using MDS data for multiple policy purposes with regards to data quality.

In the meantime, these findings should caution policy makers from using these QIs as the basis for decision making, such as with pay-for-performance initiatives. As mentioned previously, the CMS website indicates that they are considering implementing a pay-for-performance demonstration in nursing facilities, however, no details are specified on what measures would be used to assess performance and what the financial incentives would be (CMS 2007). If the four mental health-related QIs were used to reward facilities with lower poor QI values, the documentation incentives would likely be more straightforward and therefore much easier to “game” compared with the current system. For example, if facilities are given a bonus payment for having “good” QI scores for depression symptom measures, it is likely that this policy would result in a reduction in the identification and documentation of depression in nursing facility residents. Because identification and documentation are the first steps in improving care for depressed nursing facility residents, this type of strategy would likely be detrimental to the actual quality of care patients receive as opposed to working to improve it.

In conclusion, the findings from this study indicate that the documentation of mental health symptoms may be influenced by economic incentives. As a result, facilities in states that use the MDS for Medicaid reimbursement appear to perform worse on mental health measures. These mental health measures, therefore, may more correctly reflect recognition and documentation of mental health conditions instead of quality of care. As such, policy makers should avoid using these measures before further analysis and improvement.

ACKNOWLEDGMENTS

The authors wish to acknowledge Sophia Rabe-Hesketh, Ph.D., Charlene Harrington, Ph.D., and Ann Keller, Ph.D. for their helpful comments in the review of earlier drafts of this manuscript.

REFERENCES

- Arling, G., and B. Daneman. 2002. “Nursing Home Case-Mix Reimbursement in Mississippi and South Dakota.” *Health Services Research* 37 (2): 377–95.

- Arling, G., R. L. Kane, T. Lewis, and C. Mueller. 2005. "Future Development of Nursing Home Quality Indicators." *Gerontologist* 45 (2): 147–56.
- Carter, G. M., J. P. Newhouse, and D. A. Relles. 1990. "How Much Change in the Case-Mix Index Is DRG Creep?" *Journal of Health Economics* 9 (4): 411–28.
- Centers for Medicare and Medicaid Services. 2006. Available at http://www.cms.hhs.gov/NursingHomeQualityInits/10_NHMQQualityMeasures.asp website [accessed on January 10, 2007].
- . 2007. "Market Basket Data" [accessed on March 10, 2007]. Available at http://www.cms.hhs.gov/MedicareProgramRatesStats/04_MarketBasketData.asp website
- Courty, P., and G. Marschke. 2003. "Dynamics of Performance Measurement Systems." *Oxford Review of Economic Policy* 19 (2): 268–84.
- Frederikson, K., P. Tariot, and E. DeJonghe. 1996. "Minimum Data Set Plus (MDS+) Scores Compared with Scores from Five Rating Scales." *Journal of the American Geriatrics Society* 44: 305–9.
- Grabowski, D. C. 2002. "The Economic Implications of Case-Mix Medicaid Reimbursement for Nursing Home Care." *Inquiry* 39 (3): 258–78.
- Grabowski, D. C., Z. Feng, O. Intrator, and V. Mor. 2004. "Recent Trends in State Nursing Home Payment Policies." *Health Affairs Jan-Jun Suppl Web Exclusives*: W4-363–73.
- Harrington, C., J. H. Swan, B. Wellin, and W. Clemena. 2000. *1998 State Data Book on Long Term Care Program and Market Characteristics*. San Francisco: University of California, San Francisco.
- Hsia, D. C., C. A. Ahern, B. A. Ritchie, L. M. Moscoe, and W. M. Krushat. 1992. "Medicare Reimbursement Accuracy under the Prospective Payment System, 1985–1988." *Journal of the American Medical Association* 268 (7): 896–9.
- Jones, R. N., E. R. Marcantonio, and T. Rabinowitz. 2003. "Prevalence and Correlates of Recognized Depression in U.S. Nursing Homes." *Journal of the American Geriatrics Society* 51: 1401–9.
- Karon, S. L., F. Sainfort, and D. R. Zimmerman. 1999. "Stability of Nursing Home Quality Indicators over Time." *Medical Care* 37 (6): 570–9.
- Konetzka, R. T., D. Yi, E. C. Norton, and K. E. Kilpatrick. 2004. "Effects of Medicare Payment Changes on Nursing Home Staffing and Deficiencies." *Health Services Research* 39 (3): 463–88.
- Lu, M. S. 1999. "Separating the True Effects from Gaming in Incentive-Based Contracts in Health Care." *Journal of Economic and Management Strategy* 8 (3): 383–431.
- Mor, V. 2005. "Improving the Quality of Long-Term Care with Better Information." *Milbank Quarterly* 83 (3): 333–64.
- Mor, V., D. Gifford, J. Morris, and T. Moore. 2003. "Benchmarking and Quality in Residential and Nursing Homes: Lessons from the U.S." *International Journal of Geriatric Psychiatry* 18: 258–66.
- Morris, J., T. Moore, R. Jones, V. Mor, J. Angelelli, K. Berg, C. Hale, S. Morris, K. Murphy, and M. Rennison. 2003. "Validation of Long-Term and Post-Acute Quality Indicators." CMS Final Report, June 10, 2003 [accessed

- on January 10, 2007]. Available at <http://www.interrai.org/applications/qireport2.pdf>
- Office of Inspector General. 2001. "Nursing Home Resident Assessment Resource Utilization Groups." OEI-02-99-00041. Department of Health and Human Services. Available at <http://oig.hhs.gov/oei/reports/oei-02-99-00041.pdf>
- Rabe-Hesketh, S., and A. Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press.
- Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein. 2005. "Early Experience with Pay-for-Performance." *Journal of American Medical Association* 294 (14): 1788–93.
- Rowe, J. W. 2006. "Pay-for-Performance and Accountability: Related Themes in Improving Health Care." *Annals of Internal Medicine* 145: 695–9.
- Ryan, J., R. I. Stone, and C. R. Raynor. 2004. "Using Large Data Sets in Long-Term Care to Measure and Improve Quality." *Nursing Outlook* 52: 38–44.
- Sangl, J., D. Saliba, D. R. Gifford, and D. F. Hittle. 2005. "Challenges in Measuring Nursing Home and Home Health Quality—Lessons from the First National Healthcare Quality Report." *Medical Care* 43 (3 suppl): I24–32.
- Schnelle, J. F., S. Wood, E. R. Schnelle, and S. F. Simmons. 2001. "Measurement Sensitivity and the Minimum Data Set Depression Quality Indicator." *Gerontologist* 41: 401–5.
- Silverman, E., and J. Skinner. 2004. "Medicare Upcoding and Hospital Ownership." *Journal of Health Economics* 23 (2): 369–89.
- Simmons, S. F., M. P. Cadogan, G. R. Cabrera, N. R. AlSamarrai, J. S. Jorge, L. Levy-Storms, D. Osterweil, and J. F. Schnelle. 2004. "The Minimum Data Set Depression Quality Indicator: Does It Reflect Differences in Care Processes?" *Gerontologist* 44: 554–64.
- Snowden, M. 2004. "The Minimum Data Set Depression Rating Scale (MDSDRS) Lacks Reliability for Identifying Depression among Older Adults Living in Nursing Homes." *Evidence-Based Mental Health* 7: 7.
- Steinwald, B., and L. A. Dummit. 1989. "Hospital Case Mix Change: Sicker Patients or DRG Creep?" *Health Affairs* 8(2): 35–47.
- U.S. General Accounting Office. 2002a. "Elderly Individuals Could Find Significant Variation in the Availability of Medicaid Home and Community Services." Testimony before the Special Committee on Aging, U.S. Senate, 02-1131T, September 2002.
- . 2002b. "Federal Efforts to Monitor Resident Assessment Data Should Complement State Activities." Report to Congressional Requestors, 02-279, February 2002.
- Weissert, W. G., and M. C. Musliner. 1992. "Case-Mix Adjusted Nursing Home Reimbursement: A Critical Review of the Evidence." *Milbank Quarterly* 70 (3): 455–90.
- Zimmerman, D. R., S. L. Karon, G. Arling, B. R. Clark, T. Collins, R. Ross, and F. Sainfort. 1995. "Development and Testing of Nursing Home Quality Indicators." *Health Care Financing Review* 16 (4): 107–27.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:
Appendix A1. Methods for Calculating MDS Poor QI Measures.

This material is available as part of the online article from: <http://www.blackwellsynergy.com/doi/abs/10.1111/j.1475-6773.2007.00769.x> (this link will take you to the article abstract).

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