Do Trends in the Reporting of Quality Measures on the Nursing Home Compare Web Site Differ by Nursing Home Characteristics?

Jacqueline Zinn, PhD, William Spector, PhD, Lillian Hsieh, MHS, Dana B. Mukamel, PhD
Do Trends in the Reporting of Quality Measures on the Nursing Home Compare Web Site Differ by Nursing Home Characteristics?

Jacqueline Zinn, PhD,1 William Spector, PhD,2 Lillian Hsieh, MHS,3 and Dana B. Mukamel, PhD3

Purpose: This study examines the relationship between the first set of quality measures (QMs) published by the Centers for Medicare and Medicaid Services on the Nursing Home Compare Web site and five nursing home structural characteristics: ownership, chain affiliation, size, occupancy, and hospital-based versus freestanding status. Design and Methods: Using robust linear regressions, we examined the values of the QMs at first publication and their change over the first five reporting periods, in relation to facility characteristics. Results: There were significant baseline differences associated with these facility characteristics. Pain, physical restraints, and delirium exhibit a clear downward trend, with differences between the first QM reporting period and the fifth ranging from 12.7% to 46.0%. However, there were only minimal differences in trends associated with facility characteristics. This suggests that the relative position of facilities on these measures did not change much within this time period. The variation by facility type was larger for the short-stay QMs than for the long-stay measures. Implications: Those QMs that show an improvement exhibit it across all types of facilities, irrespective of initial quality levels. Although a number of alternatives may explain this positive trend, the trend itself suggests that report cards, to the extent that they are effective, are so for all facility types but only some QMs.

Key Words: Nursing Home Compare, Medicare, Consumer information, Quality measures

Nursing homes are an important component of the health care system serving older adults in the United States. About 3 million elderly and disabled Americans received care in our nation’s nearly 17,000 Medicare- and Medicaid-certified nursing homes in 2001. Slightly more than half of these individuals were long-term nursing home residents with long-term-care needs, but nearly as many had shorter stays for rehabilitation care after an acute-care episode (Nursing Home Quality Initiative, 2002).

Despite federal legislation enacted over a decade ago to improve quality, public misgivings concerning nursing home care persist, making nursing home placement a difficult decision that many patients and families would rather avoid (Harrington & Carillo, 1999; Institute of Medicine, 1986; U.S. General Accounting Office, 1999). Lack of public information about facility-specific nursing home quality has contributed to the uncertainty surrounding this decision. Making information available to the public on the performance of nursing facilities has been viewed as an important step in enabling consumers to better assess nursing facilities and to choose facilities based on information about quality (Mukamel & Spector, 2003).

In 1998, in an effort to assist consumers in evaluating the quality of nursing facilities, the Centers for Medicare and Medicaid Services (CMS) made available to consumers the Nursing Home Compare Web site (http://www.medicare.gov/NHCompare). This site provides consumers with information about Medicare- and Medicaid-certified nursing homes in the United States. It includes the quality-deficiency citations issued by state inspectors, staffing levels, and basic facility characteristics (Fermazin, Canady, Bauer, & Cooper, 2003).

In November 2002, the CMS launched the Nursing Home Quality Initiative, which introduced quality measures (QMs) to supplement the information about citations and staffing found on Nursing Home Compare. These QMs are intended to provide...
The development of the QMs was not without controversy. Although there have been a number of studies pointing to the reliability and validity of the MDS data items used to construct the measures, there have been other studies that dispute these findings (Mor, 2004). On one hand, two studies of the reliability of MDS data conducted by the Office of the Inspector General (OIG) found significant discrepancies between OIG reviewer audits and facility records (OIG, 2001a, 2001b). On the other hand, a large field-reliability trial found that 83% of MDS data elements manifest adequate interrater reliability (Mor).

The team that developed the current MDS indicators reported an association between a nursing home’s score on the pressure-ulcer-prevalence QM and pressure-ulcer-care processes as evidence of the validity of this QM (Morris, Moore, Jones, & Mor, 2002). However, several recent research studies, although small and focused in a single region, have called into question the accuracy of the QMs. For example, Bates-Jensen, Cadogan, and colleagues (2004) found no differences in nursing care in homes reporting high and low pressure-ulcer prevalence in 16 Southern California nursing homes. Chu, Schnelle, Cadogan, and Simmons (2004) found that 63% of residents with daily pain did not have moderate or serious pain reported on the MDS. In particular, residents with cognitive impairment were significantly less likely to have serious pain documented on the MDS. In addition, Cadogan, Schnelle, Yamamoto-Mitani, Cabrera, and Simmons (2004) found that facilities reporting higher prevalence of pain were more accurate than those reporting lower rates. Bates-Jensen, Schnelle, Alessi, Al-Samarai, and Levy-Storms (2004) observed high levels of residents’ being left in bed all day, at variance with what was reported in the MDS. Schnelle, Bates-Jensen, Chu, and Simmons (2004) noted inaccurate reporting (mostly underreporting) in these and other care-process areas. The potential for detection or ascertainment bias may be due to differences in the ability of clinical staff to detect patient status and symptoms, particularly in clinical areas that are harder to define or diagnose such as pain, mood, or early-stage pressure ulcers (Mor et al., 2003). Lack of formal auditing to ensure the accuracy of MDS data (with the exception of states that use MDS data as a basis for payment) was cited as a validity issue in a U.S. General Accountability Office report (GAO, 2002a).

Although conflicting results preclude a definitive judgment regarding the accuracy of QM reporting, when used under proper research conditions, it appears that the MDS can yield reliable, valid, and sensitive measures of clinical care processes, as well as patient outcomes in important clinical dimensions. However, it appears to have some shortcomings in real-world applications (Mor, 2004; Sangl, Saliba, Gifford, & Hittle, 2005). A recent article suggests that at least some of these shortcomings could be remedied. Arling, Kane, Lewis, and Mueller (2005), after critically evaluating QIs from the perspective of theory, measurement, and application, recommended a number of strategies to improve the comprehensiveness, reliability, validity, and relevance for quality assessment.

Risk-adjustment procedures also were a key area of dispute in QM development. For example, two of the three consultants hired by the National Quality Forum recommended against the use of facility-level risk adjustment at the time of publication until the approach, along with other risk-adjustment methods,
could be validated (GAO, 2002b). Consequently, the final recommendation of the GAO was that nationwide reporting be delayed until the validity of the indicators selected could be ascertained and the appropriateness of the risk-adjustment methodology was better understood.

Despite the GAO recommendation, the CMS moved forward with implementation in November 2002, citing the importance of public accountability and ongoing efforts to improve the QMs. The set of QMs included were based on the National Quality Forum recommendations and had only limited risk adjustment. The CMS advertised the introduction of the QMs in major daily newspapers to help raise awareness of the quality initiative throughout the country and promote consumer use. The Medicare Compare Web site has become one of the most popular sites on Medicare.com (National Quality Forum, 2004).

In this article, we provide an initial descriptive analysis of the first set of QMs that were published on the Medicare Compare Web site between November of 2002 and January 2004. Our objective is to offer a first look at how these reported measures vary across nursing homes by facility type and whether these measures changed over the first year of publication.

**Description of Nursing Home QMs**

The nursing home QMs are computed by the CMS from resident-assessment data (the MDS) that nursing homes routinely collect on all residents at specified intervals during their stay. The assessment records the residents’ physical and clinical conditions and abilities, as well as life-care wishes. The methodology used in calculating the QMs is described in the *Quality Measures for National Public Reporting: User’s Manual* (Abt Associates, 2002b). All of the measures are defined as rates of positive or negative health outcomes. Some have been adjusted to account, at least partially, for differences in case mix across nursing facilities. All QMs exclude residents who have missing MDS-assessment data elements needed for calculation.

The 10 measures chosen for public reporting on the Nursing Home Compare Web site from November 2002 to January 2004 covered two distinct resident populations: postacute residents, who enter a nursing home for a short stay following an acute-care hospitalization, and long-term-care residents, who enter a nursing home for a long period of time (typically until death) because they are no longer capable of living in the community. QMs were calculated for every facility in the United States. Some QMs are risk adjusted by using resident-level covariates, a facility admission profile, or both. In January 2004, the Nursing Home Compare Web site was updated with an enhanced set of 14 QMs (http://www.cms.hhs.gov/nhqi/quality). In this study we focus on the initial set of indicators that were published in the first five reporting periods.

For the chronic-care (long-stay) measures, calculations are based on any resident with a full or quarterly MDS survey in the target quarter beginning April 2002 through June 2002 (http://www.cms.hhs.gov/nhqi/quality). Because sample size affects the stability of facility scores, nursing homes with small resident numbers for a specific QM were excluded from measure calculations (Sangl et al., 2005). The facility had to have at least 30 cases within the QM category during the target time period to be included in this calculation. The six long-stay resident measures consisted of the percentage of residents in physical restraints, or with infections, pain, pressure sores, pressure sores as adjusted by the facility admission profile, or loss of ability in daily tasks:

1. The percentage of residents with loss of ability in daily tasks (Abt Associates, 2002a) measure reflects the percentage of residents who have unexpected, sudden, or rapid loss of ability to do one or more activities of daily living (ADLs). These ADLs include feeding oneself, moving from one chair to another, changing positions while in bed, and going to the bathroom alone.
2. The percentage of residents with infections (Abt Associates, 2002a) measure reports the percentage of residents who, since being admitted to the facility, have acquired a new infection.
3. The percentage of residents with pain (Abt Associates, 2002a) measure reports the percentage of residents with moderate pain daily, or extreme pain at any frequency.
4. The percentage of residents with pressure sores measure is self-explanatory.
5. The percentage of residents with pressure sores as adjusted by the facility admission profile (Abt Associates, 2002a) measure is also self-explanatory.
6. The percentage of residents in physical restraints (Center for Health System Research and Analysis, 2003) measure reports the percentage of residents in physical restraints daily.

For postacute measures, calculations were based on any resident with a 14-day MDS report in the two consecutive target quarters beginning January 2002 through June 2002. The facility had to have at least 20 cases during the target time period to be included in this calculation (http://www.cms.hhs.gov/nhqi/quality). The four postacute (short-stay) measures included the percentage of postacute residents with improved or maintained walking ability, or with pain, delirium, or delirium as adjusted by the facility admission profile:

1. The percentage of postacute residents with delirium measure is self-explanatory.
2. The percentage of postacute residents with delirium as adjusted by the facility admission profile (Abt Associates, 2002a) measure is also self-explanatory.

3. The percentage of postacute residents whose walking ability has improved, or been maintained, since admission (Abt Associates, 2002a) measure is the percentage of postacute residents who walked as well or better on Day 14 as on Day 5 of their stay. This is the only reported measure in which higher percentages are better.

4. The percentage of postacute residents with pain (as already described under long-stay measures) measure is self-explanatory.

The definition, exclusions, resident-level covariates, or facility-admission-profile adjustment and source for each of this first set of QMs are summarized in Table 1.

The CMS released the results of a preliminary analysis on changes in quality performance over the first year of Nursing Home Compare reporting, indicating changes in several measures, most notably declines in the rate of reported pain (Medical NewsService.com, 2004). Our study expands on these preliminary findings by examining whether the reported QMs recorded for the first reporting period and trends in these measures over the next four periods differ by facility characteristics that are known from prior studies to be associated with quality of nursing home care. These characteristics include the following: (a) ownership status (Aaronson, Zinn, & Rosko 1994; Angelelli, Mor, Intrator, Feng, & Zinn 2003; Arling, Nordquist, & Capitman, 1987; Cohen & Dubay, 1990; Harrington, Woolhandler, Mullan, Carrillo, & Himmelstein, 2001; Harrington, Zimmerman, Karon, Robinson, & Beutel, 2000; Mor et al., 2003; Spector, Seldon, & Cohen 1998), (b) chain affiliation (Greene & Monahan, 1981; Harrington et al., 2000, 2001), (c) size (Banaszak-Holl, Zinn, & Mor, 1995; Davis 1991; Greene & Monahan), (d) occupancy rate (Greene & Monahan; Hughes, Lapane, & Mor, 2000), and (e) hospital-based versus freestanding status (Shaughnessy, Schlenker, & Kramer, 1990).

---

Table 1. Definition of Quality Measures Reported on Nursing Home Compare, November 2002 through January 2004

<table>
<thead>
<tr>
<th>Quality Measure Definition</th>
<th>Exclusionsa</th>
<th>Resident Level Covariates</th>
<th>FAPb</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of residents with an unexpected loss of function in some ADLs</td>
<td>Comatose, end stage, hospice status, or status unknown; late-stage ADL status</td>
<td>None</td>
<td>No</td>
<td>CHSRA</td>
</tr>
<tr>
<td>% of residents with infections</td>
<td>Admission assessment; end-stage disease or hospice status (or status unknown)</td>
<td>None</td>
<td>No</td>
<td>MEGA QI</td>
</tr>
<tr>
<td>% of residents with pain</td>
<td>Admission assessment</td>
<td>Prior assessment of independent or modified independent decision making</td>
<td>No</td>
<td>MEGA QI</td>
</tr>
<tr>
<td>% of residents with pressure sores</td>
<td>Admission assessment</td>
<td>None</td>
<td>No</td>
<td>CHSRA</td>
</tr>
<tr>
<td>% of residents with pressure sores (FAP adjusted)</td>
<td>Admission assessment</td>
<td>None</td>
<td>Yes</td>
<td>CHSRA</td>
</tr>
<tr>
<td>% of residents in physical restraints</td>
<td>Admission assessment</td>
<td>None</td>
<td>No</td>
<td>CHSRA</td>
</tr>
<tr>
<td>Postacute care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of short-stay residents with delirium</td>
<td>Comatose, end-stage disease, or hospice status (or status with respect to above unknown)</td>
<td>Prior residential history preceding the current SNF stay</td>
<td>No</td>
<td>MEGA QI</td>
</tr>
<tr>
<td>% of short-stay residents with delirium (FAP adjusted)</td>
<td>Comatose, end-stage disease or hospice status (or status with respect to above unknown)</td>
<td>Prior residential history preceding the current SNF stay</td>
<td>Yes</td>
<td>MEGA QI</td>
</tr>
<tr>
<td>% of short-stay residents with moderate pain at least daily or horrible or excruciating pain at any frequency</td>
<td>None</td>
<td>None</td>
<td>No</td>
<td>MEGA QI</td>
</tr>
<tr>
<td>% of short-stay who walk as well or better on day 14 as on day 5 of stay</td>
<td>Comatose, end-stage disease, ventilator dependent, quadriplegic, paraplegic, or hospice status on day 14 of stay</td>
<td>None</td>
<td>Yes</td>
<td>MEGA QI</td>
</tr>
</tbody>
</table>

Notes: FAP = facility admission profile; ADL = activity of daily living; CHSRA = Center for Health Systems Research and Analysis of the University of Wisconsin–Madison; MEGA QI = Mega quality indicator.
aOther than missing or inconsistent data, and/or failure to trigger quality measure.
bFAP is measured at the facility level, and once calculated, it is the same for every resident in that facility. The FAP reflects the resident population admitted during a 12-month period and is intended to capture the admitting characteristics of individual facilities.
Methods

Source of Data

We downloaded data from the Nursing Home Compare Web site beginning with the first reporting period that was published in November 2002. Every 3 months, the CMS updates the data on the Web site. The last data download for this analysis was January 2004. In our data set, we have quality measures for five time periods, with each QM based on MDS data spanning the previous 3 months. Because of the lag time associated with MDS data collection and submission to Nursing Home Compare, the five time periods represent data from April 2002 to June 2002, July 2002 to September 2002, October 2002 to December 2002, January 2003 to March 2003, and April 2003 to June 2003.

Variable Measurement

We measured chain affiliation and hospital-based versus freestanding status as dichotomous variables (0, 1). We trichotomized occupancy rate as high if the facility was in the upper quartile of the occupancy distribution (95% + occupancy), medium if between the first and third quartile (78–94% occupancy), and low if in the bottom quartile of the occupancy distribution (less than 78% occupancy). We classified nursing home ownership status as for-profit, nonprofit, and public (governmentally operated). We trichotomized size as small (fewer than 60 beds, the bottom quartile of the size distribution), medium (between 60 and 124 beds, between the first and third quartiles) and large (more than 125 beds, in the upper quartile of the size distribution).

Statistical Analyses

We performed robust linear regression analyses to determine if the score at baseline and the pattern of change in these measures over the first five reporting periods were related to nursing home characteristics. We estimated separate models for each QM and for each facility characteristic of the following form (using chain affiliation as an example):

$$Q_{Mi,t} = \alpha + \beta \times t + \gamma \times Chain_{i,j} + \delta \times t \times Chain_{i,j}$$

Each model included a time variable ($t$) to capture the average trend in the QM over the 15-month period. Each model also included one of the five facility characteristics to allow us to test the hypothesis that at Time 0—that is, time of the first publication—facilities that differ by this characteristic have different QM scores (the Chain dummy variable in the equation indicates whether the facility is in a chain or not). Lastly, each model included an interaction of time and facility characteristic ($t \times Chain$). This last term allows us to test the hypothesis that the trend was different for facilities with different characteristics.

To interpret the results from this regression, note that the values for $Q_{Mi,t}$ for facilities that are not part of a chain are given by setting $Chain = 0$ and for facilities that are in a chain by setting $Chain = 1$. Therefore, $\alpha$ is the average value of $Q_{Mi,t}$ at $t = 0$, that is, at baseline for nonchain facilities, and $\beta$ is the amount that reported quality changes each time period for nonchain facilities. Here $\gamma$ is the difference in $Q_{Mi,t}$ at $t = 0$ between chain and nonchain facilities, and $\delta$ is the difference in the linear trend over time between chain and nonchain facilities. Thus, if $\gamma$ and $\delta$ are not significantly different from zero, we conclude that there are no significant differences between the reported QM for chain and nonchain facilities at baseline and over time, respectively.

Results

Nursing Home Characteristics and QM Performance at the Initial Reporting Period (November 2002)

Table 2 presents a comparison of the six long-stay resident QM scores by facility type at the time of the first publication (baseline), based on predicted values from the regressions. Within each organizational category, the differences in QM score ranged from 3% to 26%, depending on type, with an average difference across all categories of 11.8%. For example, the percentage of residents with pain is 17.6% higher in hospital-based facilities than in freestanding facilities $[(12.63 - 10.74)/10.74 = 17.6%]$. With a few exceptions, nonprofit, nonchain, smaller sized, and high-occupancy facilities reported better QM scores at the initial reporting period. Hospital-based facilities reported better scores only for physical restraints and infections. Public facilities were similar to nonprofit facilities and generally reported better scores than for-profit facilities. The results for the facility-admission-adjusted and nonadjusted pressure-sore measures were consistent except that the effect of size was lower for the risk-adjusted measure. The main exceptions occurred with pain: medium- and large-size and freestanding facilities report better QM scores and public facilities report worse results.

Table 3 presents the reported mean QM scores at baseline for the short-stay resident QMs by facility characteristics. Compared with the long-stay measures, the range of effect sizes was larger for short-stay residents (from about 5% to 62%), averaging 24.9% across all categories. The largest differences were for delirium and pain. For example, the percentage of short-stay residents with pain at baseline is 62% higher in hospital-based facilities than in freestanding facilities $[(39.2 - 24.32)/24.32 = 62%]$. For short-stay residents, the pattern of association with facility characteristics for delirium and pain
were similar. Generally, for-profit, medium- and large-sized, medium- and high-occupancy, chain, and freestanding facilities reported better outcomes. Fewer differences were shown for the walking measure, although nonchain and nonprofit facilities reported better scores.

The pattern of results for short-stay residents was different from findings for long-stay residents for some facility characteristics. For example, although public and nonprofit facilities had better scores on long-stay measures, for-profit facilities did better on short-stay measures, with the exception of the measure for walking ability. In addition, whereas nonchain facilities did better on most long-stay measures, chain facilities did better on pain for short-stay residents. Thus, we do not see a consistent effect across all measures for either ownership status or chain affiliation. However, the effects of size, occupancy, and hospital-based status were similar for the two categories of measures.

Tables 2 and 3 also report the number of facilities reporting on each QM. About 60% of facilities report the short-term QMs and 80% report the long-term QMs. QMs are not reported when a facility did not have at least 30 residents meeting the criteria for calculating the specific QM. Thus, small facilities or facilities that do not care for specific types of residents (e.g., short-term postacute patients) are excluded from the report card and from our analysis.

### Trends in QMs Over Time

Although the statistical analysis identified significant trends at the 0.01 level for all but one of the QMs, an inspection of Figure 1 shows that, for some QMs, no real trend is apparent and the overall change between the QM at the first publication and the QM at the fifth are minimal. Pain (both long term and short term), physical restraints, and delirium (with and without risk adjustment) exhibit a clear downward trend, with differences between the first QM reporting period and the fifth ranging from 12.7% to 46.0%. Thus, those measures that show a clear trend are all declining and the decline seems substantial, given the relatively short period (15 months) over which they are measured. The other measures (pressure sores, infections, walking ability, and loss in ADLs) first decline and then increase, and the overall change in the QM does not exceed 6%. Therefore, we believe it is premature to conclude that there is a trend in these latter measures.

The regression analyses allowed us to examine differences in trends by facility characteristics. Among the five QMs that exhibit a significant decline (e.g., improvement in long- and short-stay pain, physical restraints, and the two delirium measures), relatively few show a significant difference in the trend by facility characteristic. Of the 40 comparisons, only 8 were statistically significant.
These are shown in Figure 2. An inspection of these charts shows that although the differences in the rate of decline may be substantial, the trend lines never cross. Nevertheless, differences in reported measures are typically reduced by the fifth reporting period. In three cases (delirium by occupancy rate, short-stay pain by chain status, and long-stay pain by hospital status), the decline in differences associated with facility characteristic is notable. In the delirium case, facilities with low occupancy reported about a 25% higher rate at onset than did high-occupancy facilities, but by the end of the year, the difference between low- and high-occupancy levels was reduced to about 15%. For pain (short stay) at the first reporting period, nonchain facilities reported a 4% higher rate than did chain facilities, declining to 2% by the end of the year. For pain (long stay) at the first reporting period, hospital-based facilities reported a 13% higher rate than did freestanding facilities, but this declined to 6% by the end of the year.

### Discussion

The CMS began publicly reporting QMs as part of its quality initiative for nursing homes, with the expectation that the information would be useful to consumers when choosing a nursing home and for facilities in their quality-improvement efforts. This article provides an early examination of the association between initial reports of these measures and early trends by common facility characteristics. We have three important findings: (a) there are strong associations between baseline differences and organizational characteristics; (b) there are clear time trends for 5 of the 10 measures, and all are toward improved outcomes (i.e., lower prevalence rates of adverse outcomes); and (c) in general, most organizational characteristics reviewed herein are not associated with differences in trends over time, except for a few instances. Furthermore, in these cases, any differences associated with facility characteristics observed at the beginning narrow with time.

Our findings that QMs differ at baseline by organizational characteristics are consistent with prior research showing that facility attributes are associated with quality of care (Greene & Monahan, 1981; Spector et al., 1998; Spector & Takada, 1991; Zinn, Aaronson, & Rosko, 1993). In this study, for-profit and chain-affiliated facilities appear to do better on QMs (pain and delirium) for short-stay, postacute residents. Small, independent, nonprofit, high-occupancy facilities had better reported performance on QMs for long-stay residents. This suggests that some types of facilities may specialize in different aspects of nursing home care.

The trend toward improvement in five of the QMs suggests that the Medicare Compare report card, and possibly other aspects of the CMS Nursing Home Quality Initiative, may have the expected impact of improving nursing home care. However, the fact that only some measures exhibit a distinct trend raises the question of what distinguishes these measures from those that do not show a trend. Because there are a number of alternative explanations that could account for the observed positive trends on some QMs, we can only speculate on this issue. One possibility is that nursing homes that are operating with limited resources target those QMs with the highest rates, indicating the areas requiring the most attention. Indeed, the prevalence of short-term resident pain, one of the measures showing a distinct downward trend, had one of the highest prevalence rates, at over 25%. In contrast, the walking ability measure for short-term residents, with a prevalence of about 30%, did not exhibit any improvement. The lack of improvement in walking and in pressure sores may be because it is more difficult for nursing homes to affect these outcomes, especially within the short run, that is, the 15 months for which we have examined the trends. This suggests that the lack of a distinct trend toward improvement may not be due to lack of effort but rather to other factors, such as ability to affect changes in the short run. Finally, as previously noted, at least one study found no differences in nursing care in homes reporting high and low pressure ulcer prevalence (Bates-Jensen, Cadogan, et al., 2004). This suggests that the level of

### Table 3. Mean Quality Measure Scores at Baseline by Nursing Home Characteristics for Short-Stay Residents

<table>
<thead>
<tr>
<th>Nursing Home Characteristic</th>
<th>Delirium, Risk Adjusted</th>
<th>Pain</th>
<th>Walk as Well or Better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For profit</td>
<td>3.51</td>
<td>24.06</td>
<td>29.54</td>
</tr>
<tr>
<td>Nonprofit</td>
<td>4.38*</td>
<td>30.37*</td>
<td>31.59*</td>
</tr>
<tr>
<td>Public</td>
<td>4.59</td>
<td>29.96*</td>
<td>30.38</td>
</tr>
<tr>
<td>Chain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.86</td>
<td>26.82</td>
<td>31.15</td>
</tr>
<tr>
<td>Yes</td>
<td>3.74</td>
<td>25.42*</td>
<td>29.45*</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>5.22</td>
<td>35.84</td>
<td>30.57</td>
</tr>
<tr>
<td>Medium</td>
<td>3.68*</td>
<td>25.07*</td>
<td>30.18</td>
</tr>
<tr>
<td>Large</td>
<td>3.39*</td>
<td>23.40*</td>
<td>29.89</td>
</tr>
<tr>
<td>Occupancy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>4.43</td>
<td>28.91</td>
<td>30.03</td>
</tr>
<tr>
<td>Medium</td>
<td>3.73*</td>
<td>25.47*</td>
<td>30.16</td>
</tr>
<tr>
<td>High</td>
<td>3.45</td>
<td>24.80*</td>
<td>30.15</td>
</tr>
<tr>
<td>Hospital based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.57</td>
<td>24.32</td>
<td>30.09</td>
</tr>
<tr>
<td>Yes</td>
<td>5.58*</td>
<td>39.32*</td>
<td>30.45</td>
</tr>
</tbody>
</table>

*Indicates significant difference with respect to the reference category (listed first) at the p ≤ .01 level.
pressure ulcer care provided in some homes may not be good enough to make a difference.

It is also possible that some of the observed improvements reflect changes in data-reporting practices rather than actual improvements in care. The potential for detection or ascertainment bias may be due to differences in the ability of clinical staff to detect patient status and symptoms, particularly in clinical areas that are harder to define or diagnose such as pain, mood, or early-stage pressure ulcers (Mor et al., 2003). Pain in particular may be subject to reporting bias. A recent study found that pain was underassessed in facilities lacking pain specialists, suggesting that the documentation of low levels of pain may actually reflect incomplete assessment (Wu, Miller, Lapane, & Gozalo, 2003). Standards for the assessment of pain may vary as a function of differences in training in addition to underlying biological differences (Mor et al., 2003). Finally, similar to “DRG creep” (i.e., coding patients

Figure 1. Trends in quality measures.
in the clinical category that generates the most revenue) observed in hospitals after the implementation of prospective payment, positive changes in pain and delirium may reflect perverse incentives to upcode in areas that are more difficult to audit.

The current set of QMs published on the Nursing Home Compare Web site are different from the original set we report on in several respects: new measures have been added, some from the original set have been eliminated, and, for some, risk adjustment was modified. The publication of the QMs was intended to be an evolutionary process. In commenting on the recommendations of the GAO report, the CMS reiterated its commitment to continuously improve the indicators as more evidence of validity and reliability becomes available. It is prob-
ably a good thing to rotate the published QMs to avoid “teaching to the test” incentives, that is, incentives to focus only on improving the published QMs to the possible detriment of other areas of patient care. The downside of changing QMs is that we will not be able to follow trends for the eliminated QMs in the future.

The limited association between organizational characteristics and improvement trends raises concerns that the gaps in outcomes by facility type at baseline may persist. An obvious objective for report cards is to provide incentives for all facilities to improve quality, and in particular for those who lag behind. In a few instances we indeed observe that gaps are closing. For example, the differences in pain among short-stay residents between chain and non-chain facilities decreased over the period. However, most of the gaps did not close. Researchers should consider several caveats when interpreting these findings. It is not clear whether the changes observed over time are attributable to quality improvement or to changes in data reporting over time. For example, is the decline in the reported prevalence of pain attributable to better treatment of pain or to less diligence by nursing homes in documenting pain? In addition, any trend in quality improvement may have preceded, rather than have been precipitated by, QM publication. One would have to examine values of these measures for periods prior to publication to determine if Nursing Home Compare reporting is influencing observed trends. In addition, because the QMs are outcome measures, their validity depends on whether they have been appropriately adjusted for differences in case mix. If the measures are not appropriately adjusted, then nursing homes could improve their QM score by changing their admission policies toward a less frail population that may be at lower risks for these adverse outcomes. In addition, if facilities do not measure the QMs comparably, measurement bias may influence the findings. Thus, researchers must pursue further studies to address these questions in order to establish the effectiveness of the reported measures in quality improvement. Nevertheless, report cards clearly do serve an important function in focusing attention on where quality improvement is needed for facilities, regulators, and consumers.

Finally, it is important to understand how facilities that have successfully made improvements were able to make these strides. What were the most important changes in the quality-assurance process that improved QM scores? Were the efforts of the quality improvement organizations instrumental in effecting improvements? Were facilities that worked with the quality improvement organizations to improve specific QMs successful? How can we change quality-assurance programs for ADL, walking, and pressure-sores QMs where no declining trends have occurred? The answers to these questions suggest the need for in-depth evaluation studies that go beyond whether nursing home quality has improved under the Nursing Home Quality Initiative to provide insights into how any gains were achieved and how improvements could be made in the future.

References


Received November 12, 2004

Accepted June 28, 2005

Decision Editor: Linda S. Noelker, PhD