Rapid Communication

Stability and Sensitivity of Nursing Home Quality Indicators

Marilyn J. Rantz,[1,2] Lanis Hicks,[3] Gregory F. Petroski,[4] Richard W. Madsen,[5] David R. Mehr,[2]
Vicki Conn,[1] Mary Zwygart-Staffacher,[6] and Meridean Maas[7]

[1]Sinclair School of Nursing, [2]Family and Community Medicine, [3]Health Management and Informatics,
[4]Statistics, School of Medicine, University of Missouri-Columbia.
[5]Department of Statistics, University of Missouri-Columbia.
[6]Department of Nursing Systems, University of Wisconsin-Eau Claire.
[7]College of Nursing, University of Iowa, Iowa City.

THE issues of stability and sensitivity of indicators to actually detect differences in quality of care in nursing facilities are not only of concern to researchers, but also to public policy makers, who are reporting indicators to providers and now to consumers. Our research team at the University of Missouri-Columbia recently completed a study of 92 randomly selected nursing homes with a range of quality. Quality of care measurement was based on a statewide analysis of 23 quality indicators (QIs) currently used by the federal government to evaluate quality of care in nursing homes (see Table 1). The study examined the cost, staffing, and quality of care information from the nursing home Minimum Data Set (MDS) and Medicaid cost report large data sets with processes of care data collected from on-site participant observations of each facility (1). This research brief explains what we learned in our study about stability and sensitivity of the MDS QIs.

BACKGROUND OF MDS QIs

Using MDS QIs to measure quality of care in nursing facilities is not without controversy. Issues of reliability of MDS data collected from facility staff completing assessment forms, interpretation of relative risk of residents to develop the problem conditions defined by the QIs, as well as under- or over-reporting of the problem conditions, prompted a national evaluation of MDS QIs. The researchers, some of whom were developers of the MDS instrument and assessment process, concluded that there is "strong evidence" that many of the MDS QIs reviewed do "capture meaningful aspects of nursing facility performance" while others remain questionable (2). The controversy also prompted two General Accounting Office (GAO) reports; one about the need for developing better review procedures to assure data accuracy (3) and another recommending the delay of the November 2002 national public reporting of quality indicators (4). However, public reporting of MDS QIs began as scheduled.

Our research team has conducted multiple studies using MDS QIs (5–9), confirming their accuracy (10), their usefulness in quality improvement (8,9), and their sensitivity to detect quality of care differences in facilities (1). We

believe that MDS QIs can be useful to facilities to help them improve care practices, to regulators to assist them in the survey process, and to researchers interested in pursuing questions about quality of care and resident outcomes in nursing facilities. We are committed to continuing to evaluate MDS QIs, their use in practice and research, and reporting what we learn about them in our research, such as in this study of 92 facilities with a range of quality of care as measured by MDS QIs.

CLASSIFYING FACILITIES BASED ON QIs

Using statewide resident assessment information recorded on the nursing home MDS, we grouped facilities based on how well they performed on the 23 MDS-derived QIs. QI scores were interpreted using threshold ranges for good, average, and poor facility performance with thresholds established in earlier research (6,11). Because QI data are inherently multidimensional, there is no simple way to combine different QIs to obtain an overall facility quality score. Therefore, we plotted the numbers of MDS QIs each facility had in the good or poor range; we found that no facilities had all their QIs exclusively in the good or poor range. Facilities were then classified into groups by considering the numbers of MDS QIs each facility had in the good, average, or poor ranges. Our definitions allowed for a facility to have QIs in various ranges as follows: let G, A, and P denote the number of QIs in the Good, Average, and Poor range, respectively. A facility was classified as Good if $G \geq 5$ and $(G\text{-}P) \geq 2$. It was classified as Poor if $P \geq 5$ and $(P\text{-}G) \geq 5$. A facility was classified as Average if it was neither Good nor Poor and if $A \geq 15$. We randomly selected facilities from the three groups, those with mostly good performance, those with mostly average performance, and those with mostly poor performance on the QIs. We then observed care delivery processes and analyzed staffing, staff mix, and costs in each facility.

STABILITY OF QI FACILITY PERFORMANCE

When we designed the study, we made an assumption that the QIs are relatively stable for facilities. This assumption

Table 1. Minimum Data Set Quality Indicators Used in This Study

New fracture
Falls
Behavioral symptoms
Symptoms of depression
Depression without antidepressant therapy
Use of 9 or more medications
Onset cognitive impairment
Bladder or bowel incontinence
Incontinence without a toileting plan
In-dwelling catheters
Fecal impaction
Urinary tract infection
Weight loss
Tube feeding
Dehydration
Bedfast
Decline in late-loss Activities of Daily Living
Antipsychotic use
Antianxiety/hypnotic use
Hypnotic use
Daily physical restraints
Little or no activity
Pressure ulcers

was based on our prior work analyzing MDS-derived QIs (5,7), our on-site clinical experience with facilities in the state (8,9), and reports of stability from earlier work of other researchers (12). However, as the qualitative and quantitative results of this study were reviewed by our research team, a question about stability of the MDS QIs and our classification approach was raised. This led to two additional post hoc analyses: calculation of MDS QI scores and classification of the facilities into the good, average, and poor outcome groups using MDS data *at the point of observation* of the facility (there was an average of 5 months between sample selection and actual observation of each facility), and a statewide analysis of facility level MDS QI stability comparing 6 months of MDS data with a subsequent 6 months.

The first post hoc analysis revealed that only 11 of the 30 facilities classified at the time of sample selection as Group 1 (good) remained classifiable as good at the time of observation; the remaining 19 facilities "slid" into average ($n=16$) or poor ($n=3$). Within the original Group 2, 16 of the 32 classified as average continued to meet criteria for that group membership at the time of observation; of the remaining 16 facilities, 6 improved and 10 declined. Within the original Group 3, 20 of the 30 remained as poor at the time of observation; the remaining 10 improved to average. Figure 1 illustrates the changes in classification at the point of sample selection and the point of observation of each facility.

The second post hoc analysis was a statewide analysis of all facilities ($N = 486$). We used 2 years of MDS data in two annual data sets and our same classification approach. We found that only 45% of facilities were classified into the same group using two consecutive 6-month periods of MDS data (first 6 months of 2000 with last 6 months of 2000 and first 6 months of 2001 with last 6 months of 2001).

When we discovered the apparent instability in our classification approach using QIs, all qualitative and quantitative data in our study were reanalyzed using a six-group approach to the data. Three groups that were categorized the same at selection and observation (Groups 1A, 2A, and 3A, with predominately good, average, or poor QI scores) and three groups that were different at selection and observation (Groups 1B, 2B, and 3B).

Time of Sample Selection   Time of Observation

Good   30   11
6
16
Average   32   16
10
3
10
Poor   30   20
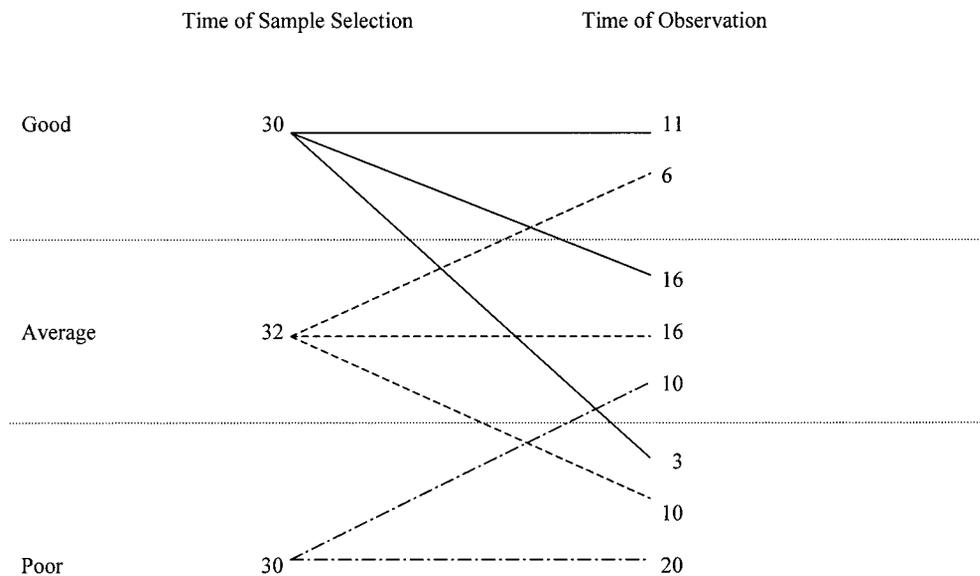
Note: There was an average of 5 months time between sample selection and actual on-site participant observation of care processes in sample facilities.

Figure 1. Changes in classification of facilities at time of sample selection and actual observation.

Table 2. MDS QIs That Discriminate Between Groups of Facilities With Good, Average, or Poor Resident Outcomes

| | | Median QI Scores | | | | |
		Group 1A Good (n = 11)	Group 2A Average (n = 16)	Group 3A Poor (n = 20)	Group Differences*	p Value
QI No.	QI Title					
2	Falls	11.9 (7.8–15)	14.2 (9.1–24.1)	17.4 (11.9–24.8)	G/AP**	.007
4	Depression	1.6 (0–11)	9.3 (4.1–27.9)	16.2 (1–33.5)	G/AP**	.002
5	Depression w/o treatment	1.6 (0–6.8)	3.8 (0.9–19.7)	8.4 (0.7–15.7)	G/P**	.011
6	Use of 9+ different medications	26 (20.3–42.9)	44.9 (26.6–57.4)	41 (27.8–57.7)	G/AP**	.020
7	Onset cognitive impairment	0 (0–12.8)	7.1 (0–12.5)	14 (3.6–22.2)	GA/P	.0003
8	Bladder/bowel incontinence	40 (9.4–57.8)	41.9 (31.3–63.8)	59.6 (48.1–78)	GA/P	.0001
9	Incontinence w/o toileting plan	29.4 (8.3–66.7)	12.9 (0–37.5)	51.6 (12.5–91.9)	A/P	.0007
12	Urinary tract infection	3.4 (0–6.7)	6.6 (4.6–11.8)	9.0 (1.6–15.8)	G/AP**	.01
14	Weight loss	3.2 (0–10)	12.4 (8.2–16.5)	12.8 (8.1–21.6)	G/AP**	.0003
16	Dehydration	0 (0–0)	0.5 (0–2.4)	0.2 (0–2.9)	G/AP**	.041
17	Bedfast residents	0 (0–5.1)	2.3 (0–7)	4.9 (1.6–12.7)	G/P**	.002
18	Decline in late-loss ADLs	6.9 (0–12.9)	12.4 (7.3–20)	16.1 (10.6–28.8)	G/AP**	.0004
26	Daily physical restraints	2.3 (0–17)	6.7 (0–17.6)	14.3 (1.6–26.3)	GA/P	.021
29	Stage 1–4 pressure ulcers	2.3 (0–6.8)	6.4 (1.6–12.9)	10.6 (6.4–18.9)	G/A/P**	.00004

Notes: *Group differences statistically significant.

**Discriminates good outcome group from average and poor outcome groups in the expected direction.

QI scores range from 0–100 with lower scores being best and higher scores being worst.

The 10th and 90th percentile scores are ( ) behind the median to display the range of scores.

MDS = minimum data set; QI = quality indicator; ADL = activities of daily living; G = Group 1A with predominately good resident outcomes as measured by MDS QIs; A = Group 2A with predominately average resident outcomes as measured by MDS QIs; P = Group 3A with predominately poor resident outcomes as measured by MDS QIs.

SENSITIVITY OF QIS TO DISCRIMINATE BETWEEN GROUPS

When the sample of facilities for the study was selected, groups were formed based on the facility QI scores for 23 QIs and their relative position within established thresholds (6,11). Recall, no facilities had all QI scores in the good, average, or poor ranges; they had scores predominately in one category. So, it was possible that facility QI scores for a particular QI could have been in any of the ranges, regardless of the group classification. After data collection, as an analytic step to confirm group selection methods and quality indicator sensitivity (that is the ability of each QI to discriminate among groups), MDS QI scores were calculated from resident assessments with effective dates within 1 year of the on-site participant observation visit. When residents had more than one assessment, the one closest to the visit date was used.

The QIs for the groups that remained classified in their group at selection and at observation, Groups 1A, 2A, 3A, were tested for significant differences. In these groups, 14 of the 23 indicators were significantly different at the 0.05 level. Next, pair-wise comparisons were performed to see which of the three groups were different. As shown in Table 2, 10 of the 23 QIs used in this study were able to differentiate the group with good resident outcomes (Group 1A) from the other outcome groups. The median scores and 10th and 90th percentile ranges are in the expected direction to discriminate quality of care. These include falls, depression, depression without treatment, use of 9 or more medications, urinary tract infection, weight loss, dehydration, bedfast residents, decline in late-loss activities of daily living, and stage 1–4 pressure ulcers. Because the select 10 QIs differentiate the groups in the expected direction, they are likely to be the ones that are discriminating or most sensitive to detecting differences in quality of care.

IMPLICATIONS

Sensitivity of MDS QIs as quality measures is of interest to researchers (5,8,13–15) and now consumers, as policy makers make some available for public reporting (16). An important implication of our study is identifying specific QIs that can apparently serve as sensitive measures to discriminate among facilities with good, average, and poor resident outcomes. Ten of the 23 QIs that are currently in use in nursing facilities appear to be most sensitive and should be considered when selecting outcome measures for other studies about quality of care in nursing homes. While the other MDS QIs measure aspects of quality of care, these 10 appear most promising for detecting quality differences, and they represent a variety of aspects about quality of care. A variety of multiple measures are necessary to use in research about quality of care in nursing homes to deal effectively with the multidimensional facets of quality, as advised by Spector and Mukamel (15), Sainfort and colleagues (17), and Zimmerman and colleagues (18).

Another important implication of our study is for researchers to consider stability of MDS QIs when designing studies that will use them. Our recommendation is that, just as we did in this study's post hoc statewide analysis, future studies should examine consecutive 6 months of MDS QI data to check for stability, and require that facilities be consistently classified in their outcome groups for at least two consecutive 6-month periods before classifying into outcome groups. Explanations for the fluctuations between outcome classification groups using MDS QIs may be that quality of care does rapidly change within facilities, or it may be that continuous resident turnover within facilities constantly causes enough fluctuations in the scores that classification of facilities using thresholds is affected. For example, since the group

classification is determined by having QI scores above or below certain thresholds, a small change in one or two QIs that are near the threshold can cause a change in facility classification. Further research is needed to determine if MDS QIs are detecting rapid quality changes in facilities or if there is something else affecting the measures.

A positive implication about sensitivity is that the MDS QIs that differed significantly across the facilities with good, average, and poor quality outcomes as measured by MDS QIs reflect quality of care issues that are amenable to change. All of these indicators can be affected by changes in the clinical processes of care used in the facility. Pressure ulcer prevention and treatment, restraint reduction practices, restorative care, improving nutrition and hydration, treating depression, reducing routine medication use, and fall prevention programs are effective clinical practice changes that can be made by nursing home staff to improve resident outcomes and facility QI performance. Poor quality outcomes in nursing facilities are not inevitably linked with the challenges of caring for frail elders. Making changes in the way that staff care for residents can improve care and subsequently improve resident outcomes.

REFERENCES

1. Rantz MJ, Hicks L, Grando V, et al. Nursing home quality, cost, staffing, and staff-mix. *Gerontologist.* In press.

2. Morris JN, Moore T, Jones R, et al. Validation of long-term and post-acute care quality indicators. Executive summary of report to CMS, Contract # 500-95-0062/TO#2; 2002:8.

3. GAO. *Nursing Homes: Federal Efforts to Monitor Resident Assessment Data Should Complement State Activities.* GAO-02-279. February 2002.

4. GAO. *Nursing Homes: Public Reporting of Quality Indicators Has Merit, But National Implementation Is Premature.* GAO-03-187. October 2002.

5. Rantz MJ, Mehr D, Conn V, et al. Assessing the quality of nursing home care: the foundation for improving resident outcomes. *J Nurs Care Qual.* 1996;10:1–9.

6. Rantz MJ, Petroski GF, Madsen RW, et al. Setting thresholds for MDS quality indicators for nursing home quality improvement reports. *Jt Commssn J Qual Improv.* 1997;23:602–611.

7. Rantz MJ, Popejoy L, Mehr D, et al. Verifying nursing home care quality using minimum data set quality indicators and other quality measures. *J Nurs Care Qual.* 1997;12:54–62.

8. Rantz MJ, Popejoy L, Petroski GF, et al. Randomized clinical trial of a quality improvement intervention in nursing homes. *Gerontologist.* 2001;41:525–538.

9. Rantz MJ, Vogelsmeier A, Manion P, et al. Statewide strategy to improve quality of care in nursing facilities. *Gerontologist.* 2003;43: 248–258.

10. Karon SL, Roberts L. Quality indicator validation studies: Final report. Center for Health Systems Research and Analysis UW-Madison. Submitted to the University of Missouri-Columbia and the Missouri Division of Aging; 1999.

11. Rantz MJ, Petroski G, Madsen R, et al. Setting thresholds for quality indicators derived from MDS data for nursing home quality improvement reports: an update. *Jt Commssn J Qual Improv.* 2000;26: 101–110.

12. Karon S, Sainfort F, Zimmerman DR. Stability of nursing home quality indicators over time. *Med Care.* 1999;37:570–579.

13. Arling G, Karon SL, Sainfort F, Zimmerman DR, Ross R. Risk adjustment of nursing home quality indicators. *Gerontologist.* 1997;37: 757–766.

14. Mukamel DB. Risk-adjusted outcome measures and quality of care in nursing homes. *Med Care.* 1997;35:367–385.

15. Spector WD, Mukamel DB. Using outcomes to make inferences about nursing home quality. *Eval Health Prof.* 1998;21:291–315.

16. Crecelius C, Wilson KM. Making (up) the grade: what you need to know about nursing home "report cards." *Caring for the Ages.* 2002;3: 3,11,26.

17. Sainfort F, Ramsay JD, Monato H Jr. Conceptual and methodological sources of variation in the measurement of nursing facility quality: an evaluation of 24 models and an empirical study. *Med Care Res Rev.* 1995;52:60–87.

18. Zimmerman DR, Karon SL, Arling G, et al. Development and testing of nursing home quality indicators. *Health Care Financ Rev.* 1995;16: 107–127.