
Testing Consumers' Comprehension of Quality Measures Using Alternative Reporting Formats

Margaret Gerteis, Ph.D., Jessie S. Gerteis, David Newman, J.D., Ph.D., and Christopher Koepke, Ph.D.

CMS has publicly reported nursing home quality measures since 2002, but research has shown that many users do not understand them. Alternative visual displays may improve comprehension. We developed seven reporting templates in different formats, including bar graphs like those displayed on the CMS Nursing Home Compare Web site www.medicare.gov, and tested them with 90 individuals age 45-75, using structured protocols. Tests of significance were conducted, and statistically significant findings identified. Fewer than one-half the respondents accurately interpreted bar graphs as currently displayed on the Nursing Home Compare Web site. Respondents made fewest errors on templates using words to characterize performance as better, average, or worse.

BACKGROUND

In 2002, CMS launched the Nursing Home Quality Initiative (NHQI), one of several national quality initiatives designed to (1) help people with Medicare (and those who assist them) make better decisions regarding their care, (2) create incentives for improvement by making health care providers publicly accountable for the quality of care they deliver, and (3) help guide and inform those providers' improvement efforts. A major feature of this

and subsequent national quality initiatives has been public reporting of quality measures on CMS' consumer-oriented Web site, www.medicare.gov.

CMS reports quality measures for long- and short-stay nursing home residents based on data derived from the minimum data set (MDS), a standardized periodic assessment of the status and functioning of nursing home residents.¹ Web site users can search for nursing homes by geographic area and compare the performance of the facilities they select on specific quality measures. Quality measures are reported as a percentage of residents in a given facility in a given status, along with State and national averages. Comparative information about multiple nursing homes is displayed through the use of bar graphs (one set of comparative bar graphs for each selected measure). Quality measures for individual nursing homes are also displayed in tabular format. At the time of this study, CMS reported five quality measures for long-stay residents, all of which reflected potentially preventable negative outcomes: (1) percent of residents with loss of ability in basic daily tasks; (2) percent of residents with pressure sores; (3) percent of residents with pain; (4) percent of residents in physical restraints; and (5) percent of residents with infections.²

For consumers to use the publicly reported quality measures to inform their health care decisions, the first requirement

Margaret Gerteis is with Mathematica Policy Research, Inc. Jessie S. Gerteis is with the Boston Medical Center. David Newman is with Abt Associates. Christopher Koepke is with the Centers for Medicare & Medicaid Services (CMS). The research in this article was supported by CMS under Contract Number 500-00-0037(TO3) with BearingPoint, Inc. The statements expressed in this article are those of the authors and do not necessarily reflect the views or policies of Mathematica Policy Research, Inc., Boston Medical Center, Abt Associates, CMS, or BearingPoint.

¹ CMS requires the use of the MDS to collect information about each nursing home resident for use in comprehensive planning for resident care.

² CMS currently reports 12 quality measures for long-stay residents and 5 measures for short-stay residents.

is that they understand the data as reported. However, studies commissioned by CMS to inform the quality reporting initiatives strongly suggested that consumers would have difficulty understanding or using clinical measures of quality, as they relate to a facility's performance, without assistance (Barents Group, 2001, 2003). Researchers also found that older or less-educated consumers have particular difficulty understanding displays of health-care information (Schapira, Nattinger, and McHorney, 2001; Hibbard and Peters, 2003). Prior Web site testing sponsored by CMS also found that consumers cannot easily access information that requires scrolling through many different displays on long Web pages, and that they may miss information not readily visible on the computer screen (Vaiana and McGlynn, 2002; Barents Group, 2002; BearingPoint, 2003a,b). CMS' earlier consumer testing of reporting for the Consumer Assessment of Health Plans (CAHPS®) data also suggested that consumers had difficulty interpreting bar charts (Goldstein and Fyock, 2001). These and related research findings raised questions about consumers' ability to access, understand, and use comparative information as it is currently displayed in multiple bar graphs on the Nursing Home Compare Web site.

Although CMS' earlier research on CAHPS® displays suggested that star charts were often misinterpreted, later research exploring alternative ways of displaying data to facilitate consumer choice of health plans suggested that the use of visual cues to highlight better or worse performance could lower the cognitive effort required to interpret data displays, and thereby improve their usability (Goldstein and Fyock, 2001; Hibbard et al., 2002; Vaiana and McGlynn, 2002). Hibbard and colleagues (2002) also found that

presenting quality information in a more evaluable format increased the weight it carried in consumer decisionmaking.

Most prior research in this area focused on generic issues of reporting quality information, risk communication for medical decisionmaking, or reporting of information to facilitate health plan choice. No prior research directly compared alternative templates for reporting facility-specific quality performance, nor had research focused on the display of nursing home quality data. The research team therefore sought to explore how well consumers could interpret nursing home quality measures as currently displayed, and whether alternative displays demonstrably affected their ability to understand and interpret the data.

Building on prior research, we wished to explore (1) whether visual cues would help consumers interpret the displays of nursing home measures, (2) whether reporting formats other than the bar charts commonly used on CMS' Compare Web sites would enhance comprehension, and (3) to what extent consumers' self-reported preferences for one format or another correlated with accurate interpretation. We hypothesized that visual cues would facilitate interpretation. We had no explicit hypotheses as to whether alternative formats would be more or less problematic than bar charts or about the extent to which stated preferences would correlate with accurate interpretation.

The research reported here, conducted during fall and winter 2003-2004, was intended to evaluate alternative formats for consumer reports of nursing home quality measures for possible use on the Nursing Home Compare feature of CMS' consumer Web site.

RESEARCH GOALS AND OBJECTIVES

Several factors must be taken into account to determine which reporting formats best support the cognitive functions associated with interpreting reports of nursing home quality data. First, whether the display is intended to facilitate browsing (that is, a quick review of many different nursing homes across many different measures) or a more detailed review of a single home's performance across several different measures. Second, whether the data will be displayed in numeric formats that display actual performance scores on a given measure, in formats (such as bar graphs) that display quantitative performance measures graphically, or in evaluative formats that use symbols to represent better or worse performance in comparison to an identified benchmark. Third, whether information is framed positively (with higher numbers reflecting good performance) or negatively, as most nursing home quality measures are currently reported.

For the purposes of this research, the research team focused on the following:

- The browsing function, since this is an activity that Web site users would likely first engage in, when searching for a nursing home.
- Comparing the existing bar graph display format against both numeric and evaluative formats (and, within each type, testing two or three alternative displays).
- Testing the five quality measures for long-stay residents reported in 2003 as they are currently framed (rather than testing alternative positive frames).

We then set out to explore the following questions:

- Does format discernibly affect accuracy or ease of interpretation or consumer preferences?

- If so, which formats lend themselves to the easiest and most accurate interpretation of the data?
- Which formats do respondents subjectively prefer? And, are the preferred templates the same or different than the ones that best promote accuracy?
- Does the negative direction or negative framing of quality measures appear to affect understanding?

METHODOLOGY

Independent Variable—Templates for Testing

In developing alternative formats for testing, we consulted with CMS staff who developed the quality measures, designed the Nursing Home Compare Web site, and had extensive experience developing and testing the various comparison tools offered on www.medicare.gov. We also spoke with a CMS consultant on the design of materials for beneficiaries, and on the use of visual cues to convey comparative performance.

We developed three categories, or sets, of templates to be used in testing alternative formats:

- Evaluative templates using symbols or words to depict performance that varied from the State average by at least one standard deviation.
- Numeric templates displaying percentages in a table.
- Graphic templates, using bar graphs similar to those currently displayed on Nursing Home Compare.

Within each category, we also developed two or three alternative display formats, creating seven templates in all (Table 1). Although the display formats were designed to mimic those that might be found on the Nursing Home Compare Web site,

Table 1
Templates Developed for Testing Alternative Formats for Reporting Nursing Home Quality Measures

| Template | Description |
|-------------------------------------|--|
| Evaluative Table with Stars | Uses one, two, and three stars to indicate better, about average, and worse performance for individual nursing homes in comparison to the State average. |
| Evaluative Table with 3 Symbols | Uses three different symbols (stars, approximation signs, and Xs) to indicate better, about average, and worse performance (respectively) for individual nursing homes in comparison to the State average. |
| Evaluative Table with Words | Uses the words better, average, and worse to indicate individual nursing home performance in comparison to the State average. |
| Numeric Table with Percentages Only | Shows the actual percentage of residents at each nursing home with the characteristic or condition reported in each quality measure. |
| Numeric Table with Stars | Shows the percentages, as in the numeric table with percentages only, but also includes stars to indicate individual nursing home performance that is better than the State average. |
| Standard Bar Graph | Based on the bar graphs currently shown on the Nursing Home Compare Web site, shows State and national average as bars at the top, differentiated by color from bars indicating individual nursing home performance. |
| Bar Graph with Line | Similar to standard bar graph, except that State average is displayed as a vertical line cutting across bars, rather than as a separate bar. National average is not displayed. |

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

we used black and white paper mockups for testing purposes.

Each template included the five nursing home quality measures for long-stay residents reported on the Nursing Home Compare Web site at the time these templates were developed (August 2003), and data for 10 nursing homes. We designed the evaluative and the numeric templates such that data for all five measures and all 10 nursing homes could be viewed on a single page (Figures 1 and 2). However, the use of bar graphs does not permit data for all measures and all nursing homes to be displayed on a single page. Each graphic template therefore consisted of a set of five separate graphs, with each quality measure on its own page, and with each graph displaying the 10 homes' performance on a single quality measure (Figure 3).

In creating the templates, we used real nursing home names and real data, so that the variation among the nursing homes depicted would be realistic. However, the names and the data were drawn from

different markets and from markets other than those in which testing took place. Researchers informed respondents that the data displayed were for testing purposes only, did not reflect nursing homes in their area, and should not be used to evaluate any nursing home's quality or performance. The same nursing home names appeared on all templates, and the same data across all homes were used on all templates. However, the order of the data was systematically varied for each template, so that the answers derived from one testing activity could not be used to inform subsequent activities.

Testing Protocols

The research protocols used one-on-one in-person interviews structured to combine an experimental design with qualitative insight. The experimental design allowed for direct quantitative comparison of outcomes related to respondents' accurate assessment of the displays. The qualitative

component explored why respondents made errors when using the displays.

We developed structured instruments for testing and conducted two rounds of pretesting to identify areas of confusion and make needed adjustments to ensure that the protocol, instruments, and materials worked as intended. All adjustments were made prior to data collection.

Interview protocols began with the researcher briefly introducing the task at hand and presenting a hypothetical scenario that established a context for reviewing the templates. Respondents were asked to imagine they had an elderly aunt living in another State who was soon to be discharged from the hospital after suffering a stroke and would require nursing home placement. As this aunt's closest living relative, the respondent's task was to help her select a nursing home. The reported quality measures, as displayed in the testing materials, could help them determine which nursing homes to visit, recognizing that they would not have enough time to visit more than a few.

We structured the protocols such that all respondents would review all seven templates, with the two or three templates within a given category viewed together. However, we systematically varied the order in which subjects reviewed templates, both between format categories and within format categories, to mitigate possible order effects.³ Pretesting revealed that respondents found it confusing to switch between the one-page displays and the five-page bar graph displays. To reduce the number of times this type of transition was necessary, we always showed respondents the graphic templates either first or last. Testing protocols were uniquely ordered for each interview.

³ However, the small cell sizes did not permit a systematic analysis of order effect.

Interview protocols combined open-ended questions to elicit respondents' subjective understanding of each template (e.g., "What is this table showing you?") with closed-ended questions that probed their comprehension of each template. Two closed-ended comprehension questions, written to be comparable across all templates, were asked about each template: (1) a warmup question that required simply reading the information displayed (e.g., "Which nursing home has the highest percentage of residents with pressure sores?") and (2) a question that required some interpretation of the data displayed in the template (e.g., "Which nursing home(s) are performing better than the State average on the measure 'Percentage of residents with infections?'").

Recruitment

Professional research facilities, using screeners that the research team developed, recruited 90 individuals from the Boston, Massachusetts, and McLean, Virginia, metropolitan areas to participate in testing. Respondents were between the age of 45 and 75, the range determined to be representative of the family caregiver population to whom materials for the nursing home quality reporting are targeted or directed. Respondents reflected a mix of ethnic and racial backgrounds, sexes, and education levels (Table 2). Because the goal of testing was to determine which templates would be easiest to understand for all potential Web site users, prior experience with nursing home placement was not a criterion for participation.

Data Collection

We used trained interviewers to conduct interviews in BearingPoint, Inc., offices in both metropolitan areas during fall and

Table 2
Number of Respondents, by Characteristics and Testing Location

| Characteristic | Number of Respondents |
|---------------------------|-----------------------|
| Sex | |
| Male | 46 |
| Female | 44 |
| Age | |
| 45-64 Years | 56 |
| 65-75 Years | 34 |
| Ethnicity | |
| White | 41 |
| African-American | 30 |
| Asian | 6 |
| Hispanic | 11 |
| Other | 2 |
| Education | |
| High School Graduate Only | 18 |
| Some College | 24 |
| College Graduate | 16 |
| Post-Graduate | 32 |
| Testing Location | |
| Boston, Massachusetts | 27 |
| McLean, Virginia | 63 |
| Total | 90 |

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

winter 2003-2004. The interviewers followed the structured protocols previously described. After introducing the task at hand and giving the respondent the first template to review, the interviewer left the room for approximately 5 minutes to allow respondents time to review the template on their own. For each subsequent template, interviewers allowed respondents a few minutes for review before asking questions, but did not leave the room. However, when switching from bar graphs to tables, or vice versa, the interviewer once again left the room both to provide a cue about the change in format and to allow more time for adjustment. Each interview lasted from 45 minutes to 1 hour.

Interviewers recorded responses to each question (including verbatim responses to open-ended questions or probes) on a protocol sheet uniquely tailored to that interview. Data from testing were then entered into a database.

Outcome Measures

Comprehension and Accuracy

Answers to the closed-ended, comprehension questions were used to compare accuracy of interpretation across templates. Answers to these questions were coded to indicate full errors (completely wrong answers), partial errors (answers that contained both correct and incorrect information, or that were missing some correct information), and entirely correct answers. Overall scores were then calculated and compared, based on the sum of errors from the two comprehension questions per template. Thus, for each template, there were 180 possible errors that could have been coded (2 questions x 90 respondents), and a total of 1,260 possible errors (180 errors per template x 7 templates) overall.

Sources of Errors

Interviewers recorded responses to all open-ended questions. They also asked respondents to explain their answers to closed-ended questions, so that any misinterpretations would be revealed. Verbatim responses to these probes were entered into the database for all full and partial errors, so that the sources of errors for each template could be determined. When the source of an error was clear to the interviewer, even if it was not clear to the participant, this was also recorded. For example, if a respondent consistently reported that the nursing home with the highest percentage was performing the best, it was noted that the respondent appeared to be confused by the negative framing of the measures.

Preferences and Ease of Interpretation

After reviewing each of the templates from the first category (graphic, numeric, or evaluative) one by one, respondents were asked to review all of the templates from that category together and to indicate which one they preferred and which they found easiest to use. This process was repeated for each category of templates as they were tested. After reviewing all seven templates from each of the three categories, we asked respondents to select the one that they preferred overall, the one they preferred least, and the one they found easiest to use.

Data Analysis

We analyzed the relationship of each template (our independent variable) to each of the identified outcome measures of interest to the research team: the number (and type) of errors made when interpreting the template, the number who preferred that template overall, and the number who selected that template as the easiest to use and understand. We conducted tests of significance, where feasible, for all of the results reported here, and we identify all statistically significant findings, as appropriate.

The data were analyzed using the Stata 7 (StataCorp LP, 2001) statistical software program and Microsoft® Excel® in accordance with our analysis plan. This plan called for cross tabulations of the data to elicit contingency tables that contained pertinent frequency information. We conducted one variable chi-square tests for individual variables comparing the observed frequency to an expected frequency, where the expected frequency reflected either an equal distribution of respondents across templates or an equal rate endogenously derived from the data. While the sample

size and distribution of values in the contingency tables were more than adequate for chi square tests, the sample size of 90 (stratified into seven template types) limited the formal statistical analysis.

STUDY LIMITATIONS

For budgetary reasons, this study was limited to 90 respondents. While this sample was sufficient to demonstrate significant differences among alternative templates, it was not large enough to explore how factors such as respondents' age, education, or geographic location may have affected their ability to use or interpret the data displays. As the study was limited to two geographical sites, observations may not be generalizable to other geographic areas.

This study used the nursing home quality measures as then reported on the CMS Nursing Home Compare Web site, all of which reflected negative outcomes. We did not test the effects of alternative framing, and it is therefore not clear how the templates tested would fare under similar circumstances if the measures were positively framed.

This study focused on only a few templates designed to represent three types of reporting formats. The results reported here cannot predict whether consumers would respond similarly to alternative designs within the same genres.

The study tested only alternative data displays, with limited explanatory text. Although other studies (Barents Group, 2002; BearingPoint, 2003b) suggest that Web site users often do not read the narrative that accompanies the visual displays on other Compare Web sites, that text can offer more detailed explanations about the quality measures and the data displays that may affect consumers' comprehension.

Although the study team pretested the study protocols and testing materials and made adjustments prior to data collection, some of the errors reported here may reflect respondents' misunderstanding of the interview questions, rather than misinterpretation of the data displays.

RESEARCH FINDINGS

Comprehension and Accuracy

Respondents made the fewest errors when using the evaluative templates and the most errors when using the graphic templates. Of the possible 180 errors per template, the "Evaluative Table with Words" (Figure 1) elicited the fewest total errors (with only 12), followed by the "Evaluative Tables with Stars" (14 errors) and the "Evaluative Table with 3 Symbols" (22 errors). The two graphic templates elicited the most errors, with the "Standard Bar Graph" eliciting far more total errors (54) than any other template. Respondents made more interpretive errors on the numeric templates than on the evaluative ones, but fewer than on the graphic ones (Table 3).

The most common errors respondents made were the following:

- **Incomplete Answer**—This often occurred when the question required respondents to identify several nursing homes to answer the question correctly.
- **Confused by Negative Direction**—Many respondents were confused by the fact that lower percentages (on the graphic and numeric templates) meant better performance, and vice-versa. Almost one-fifth of all errors were attributable to this confusion or misinterpretation.
- **Looked at Wrong Measure**—Fourteen percent of the errors occurred because respondents looked at the wrong measure when answering a specific question. Nine of the 25 "looked at wrong measure" errors occurred when using the graphic templates, because they required consulting several different pages. Respondents often initially missed the label indicating which quality measure was being displayed on the graphic templates, but most were able to differentiate among five different graphs after examining them more closely.

Findings are similar when the data are examined by the number of respondents who correctly interpreted each template (Table 4). Nearly 90 percent of respondents were able to correctly interpret the

Table 3
Full, Partial, and Total Errors in Interpretation, by Template

| Template | Full Errors | Partial Errors | Total Errors ¹ | Percent of All Errors |
|-------------------------------------|-------------|----------------|---------------------------|-----------------------|
| Evaluative Table with Stars | 5 | 9 | 14 | 7.7 |
| Evaluative Table with 3 Symbols | 14 | 8 | 22 | 12.1 |
| Evaluative Table with Words | 7 | 5 | 12 | 6.6 |
| Numeric Table with Percentages Only | 8 | 16 | 24 | 13.2 |
| Numeric Table with Stars | 11 | 14 | 25 | 13.7 |
| Standard Bar Graph | 19 | 35 | 54 | 29.7 |
| Bar Graph with Line | 17 | 14 | 31 | 17.0 |

¹ $n=182$ total errors across all templates. The observed distribution of errors is significantly different from an equal expectation. Chi-square (= 45.0) significant at 0.01 level.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

Table 4
Number and Percent of Respondents
Correctly Interpreting Each Template

| Template | Number | Percent |
|-------------------------------------|--------|---------|
| Evaluative Table with Stars | 77 | 86 |
| Evaluative Table with 3 Symbols | 68 | 76 |
| Evaluative Table with Words | 80 | 89 |
| Numeric Table with Percentages Only | 68 | 76 |
| Numeric Table with Stars | 66 | 73 |
| Standard Bar Graph | 42 | 47 |
| Bar Graph with Line | 65 | 72 |

NOTES: $n=90$ respondents who were asked to evaluate each template. The observed distribution of errors is significantly different from an equal expectation. Chi Square ($=17.11$) significant at <0.01 level.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

“Evaluative Table with Words” (89 percent) and the “Evaluative Table with Stars” (86 percent), a significantly higher proportion than for any of the other templates tested. Notably, only 47 percent of all respondents correctly interpreted the “Standard Bar Graph,” the lowest number for any of the templates tested.

Respondent Preferences

The “Evaluative Table with Words” (Figure 1), “Numeric Table with Stars” (Figure 2), and “Evaluative Table with Stars” were the three templates most often selected as preferred overall (Table 5). Forty-one respondents (46 percent of the sample) chose one of the evaluative tables as their preferred template. Numeric templates were the next most preferred type, with a total of 29 respondents (32 percent of the sample) choosing one of these two tables. Graphic templates were least likely to be chosen as a preferred template, selected by a total of 20 (22 percent of the sample).

The “Evaluative Table with 3 Symbols” and the “Standard Bar Graph” were

Table 5
Number and Percent of Respondents
Preferring Each Template

| Template | Number | Percent |
|-------------------------------------|--------|---------|
| Evaluative Table with Stars | 17 | 19 |
| Evaluative Table with 3 Symbols | 5 | 6 |
| Evaluative Table with Words | 19 | 21 |
| Numeric Table with Percentages Only | 10 | 11 |
| Numeric Table with Stars | 19 | 21 |
| Standard Bar Graph | 6 | 7 |
| Bar Graph with Line | 14 | 16 |
| Total | 90 | 100 |

NOTES: Data shown are based on responses to the following question: “Among all of the tables and charts you looked at today, which one do you like the best?” The observed distribution is significantly different from an equal expectation. Chi-square ($=16.4$); significant at 0.05 level.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

Table 6
Number and Percent of Respondents Who
Chose Each Template as the Easiest to Use

| Template | Number | Percent |
|-------------------------------------|--------|---------|
| Evaluative Table with Stars | 20 | 22 |
| Evaluative Table with 3 Symbols | 10 | 11 |
| Evaluative Table with Words | 27 | 30 |
| Numeric Table with Percentages Only | 3 | 3 |
| Numeric Table with Stars | 12 | 13 |
| Standard Bar Graph | 5 | 6 |
| Bar Graph with Line | 13 | 14 |
| Total | 90 | 100 |

NOTES: Data shown are based on responses to the following question: “Among all of the tables and charts you looked at today, which one do you think is the easiest to use and understand?” The observed distribution is significantly different from an equal expectation. Chi-square ($=30.3$) significant at 0.01 level.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

the templates least often selected as an overall favorite.

Figure 1
Evaluative Table with Words

| | Compared to the State Average, How the Nursing Homes Performed on the Following Quality Measures... | | | | |
|--------------------------------------|--|--------------------------------------|----------------------------|---|----------------------------------|
| | Residents with Loss of Ability in Basic Daily Tasks | Residents with Pressure Sores | Residents with Pain | Residents in Physical Restraints | Residents with Infections |
| Adams Nursing Home | Average | Better | Average | Worse | Data Not Available |
| Clarke Center | Average | Better | Average | Average | Data Not Available |
| Glenwood Specialty Hospital | Average | Average | Average | Average | Better |
| Delmore Rehab and Health Ctr. | Worse | Average | Better | Better | Average |
| Frances Bird HC | Average | Better | Better | Average | Average |
| Isolde Geriatric Center | Better | Worse | Better | Worse | Average |
| Kelly Residence | Data Not Available | Average | Better | Average | Better |
| Mary Magdelene NH | Worse | Average | Worse | Average | Average |
| Thomas HCC | Data Not Available | Average | Average | Average | Better |
| Village Center Nursing Home | Average | Average | Average | Average | Average |

NOTES: This figure represents the authors' mockup of one of the templates used in consumer testing. It does not depict the actual performance of real nursing homes. The facility names shown here are fictional and are not the same as those used in testing. Otherwise, the figure has not been edited.
SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

Reported Ease Of Use

Respondents were most likely to choose the “Evaluative Table with Words” (Figure 1) or the “Evaluative Table with Stars” (30 and 22 percent, respectively) as the easiest template to use and understand (Table 6). The “Bar Graph with Line” (Figure 3), “Numeric Table with Stars”

and “Evaluative Table with 3 Symbols” all fared about the same on this question (14, 13, and 11 percent, respectively).

Respondents were least likely to choose the “Numeric Table with Percentages Only” or the “Standard Bar Graph” as the easiest to use (3 and 5 percent, respectively).

Figure 2
Numeric Table with Stars

| *** Lower Percentages Mean Better Performance *** | | | | | |
|---|---|---------------------|-----------|------------------------|-----------------|
| | Percentage of Nursing Home Residents... | | | | |
| | With Loss of Ability in Basic Daily Tasks | With Pressure Sores | With Pain | In Physical Restraints | With Infections |
| STATE AVERAGE | 15% | 9% | 4% | 9% | 16% |
| Adams Nursing Home | 30% | 17% | ★ 0% | 3% | ★ 6% |
| Clarke Center | N/A | ★ 0% | ★ 0% | 15% | 12% |
| Glenwood Specialty Hospital | 11% | 7% | 11% | 26% | 10% |
| Delmore Rehab and Health Ctr. | 21% | 9% | 4% | 7% | 16% |
| Frances Bird HC | 13% | ★ 3% | 4% | 5% | ★ 5% |
| Isolde Geriatric Center | 10% | ★ 1% | 7% | 1% | 8% |
| Kelly Residence | 33% | 13% | 3% | 5% | ★ 6% |
| Mary Magdelene NH | 13% | 7% | 3% | 20% | 17% |
| Thomas HCC | ★ 6% | 7% | ★ 0% | 13% | N/A |
| Village Center Nursing Home | N/A | 12% | ★ 0% | ★ 0% | N/A |

Note:
Percentages with a star (★) are notably **BETTER** than the state average.
All others are about the **SAME** as or **WORSE** than the state average.
N/A means that data are not available for this measure at this time.

NOTES: This figure represents the authors' mockup of one of the templates used in consumer testing and has not been edited. Data shown are based on responses to the following question: "Among all of the tables and charts you looked at today, which one do you think is the easiest to use and understand?" The observed distribution is significantly different from an equal expectation. Chi-square (=30.3) significant at 0.01 level.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc., Gerteis, J.S., Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services.

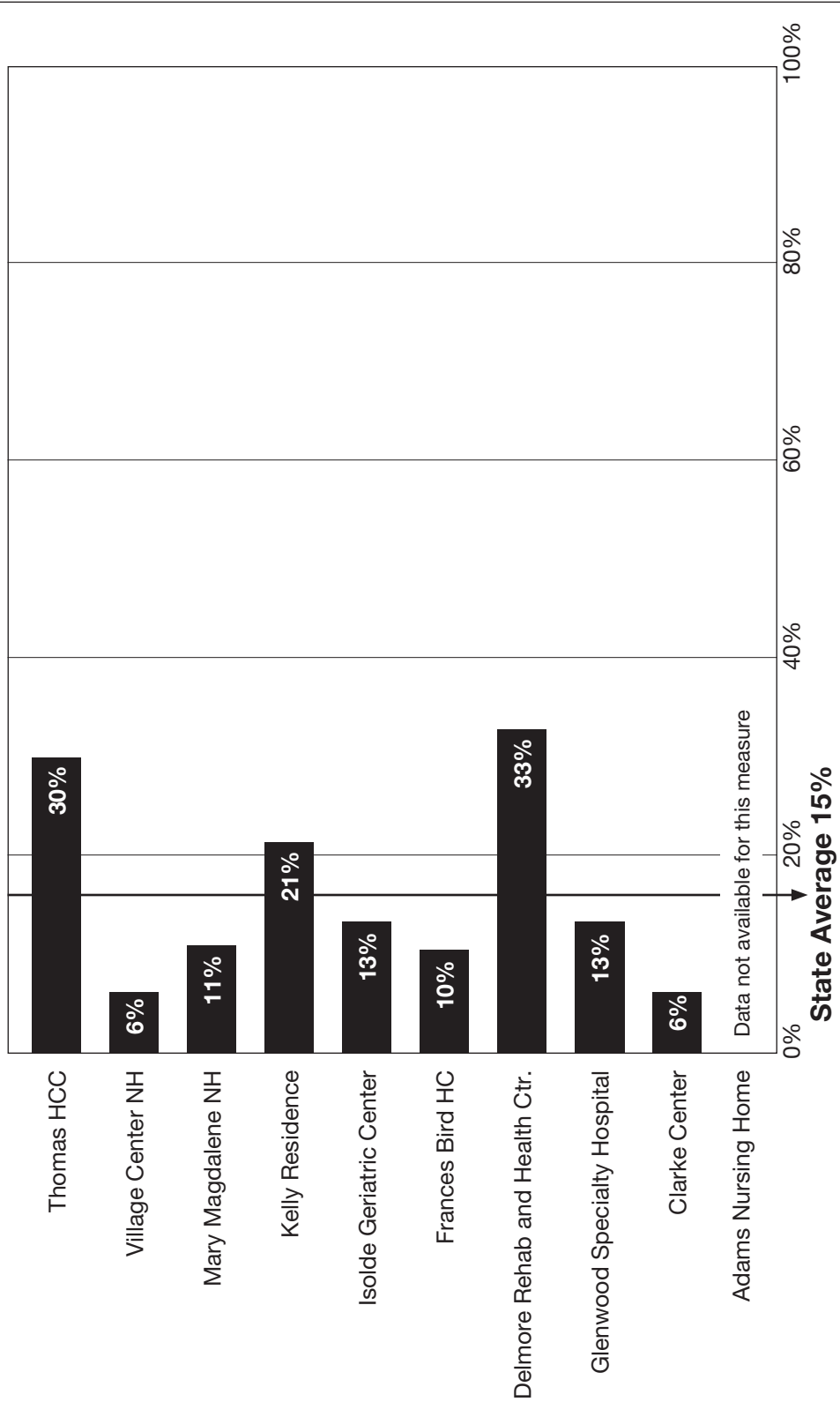
Observations

In addition to performing quantitative analyses of the data collected through the

structured protocol, the research team used qualitative techniques (team debriefing and interobserver reports) to identify common themes in respondents' answers

Figure 3
Bar Graph with Line

***** Lower Percentages Mean Better Performance *****
Percentage of Residents with Loss of Ability in Basic Daily Tasks



NOTES: This figure represents the authors' mockup of one of the templates used in consumer testing. It does not depict the actual performance of real nursing homes. The facility names shown here are fictional and are not the same as those used in testing. Otherwise, the figure has not been edited.

SOURCE: Gerteis, M., Mathematica Policy Research, Inc.; Gerteis, J.S.; Boston Medical Center, Newman, D., Abt Associates, and Koepke, C., Centers for Medicare & Medicaid Services

to open-ended questions, their observed behavior, or unsolicited comments made during the course of the interviews. The observations offered here reflect these findings, as well as those derived from the formal quantitative analyses.

Most respondents reported that they preferred to see all information on one page or in one table, for browsing purposes. Respondents pointed out that they were able to look over the evaluative and numeric tables and get an overall sense of which nursing homes were performing well and which were not. Browsing in this way was not possible with the graphic templates, because respondents could not view all of the quality measures on one page at the same time. Although approximately one-fifth of the respondents preferred one of the graphic templates, many noted that flipping back and forth among the five bar graphs, each displaying a different measure, was cumbersome and confusing. Interviewers also noted that switching among the graphs led to many errors, even when respondents reported that the graphs were easy to use.

Many respondents found the negative direction of the measures to be confusing. Although a label prominently displayed at the top of the bar graphs and numeric tables indicated “Lower Percentages Mean Better Performance,” a number of respondents mistakenly assumed that higher percentages were better. Several read and understood the label initially but forgot it when they used the data to answer questions. Moreover, participants were often unaware they were interpreting the data incorrectly, sometimes remarking that the table or graph in question was easy to use. Some explained that it was difficult to think of lower numbers as signifying better performance because it was counterintuitive and unfamiliar to them.

Some respondents who interpreted higher numbers to mean better performance assumed that the purpose of the measures was to highlight the nursing home’s expertise or capability to handle residents with identified characteristics. Some respondents thought the quality measures referred to strengths and capabilities of the nursing homes, rather than to preventable adverse outcomes. For example, if the nursing home had a high percentage of residents with pressure sores, some respondents thought this meant the home must be particularly good at dealing with pressure sores, thus attracting a higher proportion of residents with this condition. It was unclear whether these respondents interpreted higher percentages to mean good performance because of the way they understood the measures, or whether they interpreted the measures in a positive light because they assumed higher numbers meant better performance.

A number of respondents did not understand how the percentages shown related to the performance of a given nursing home on a quality measure. Regardless of how the data were displayed, some respondents thought that they reflected ratings of nursing home staff and care, rather than a percentage of residents with a certain condition. Others struggled to understand the percentages themselves, asking questions such as: “Percentage of what? I don’t know how to interpret these numbers.”

Many respondents expressed concern about nursing homes for which data were not available for one or more measures, and most reported that they would try to avoid these homes. Many respondents thought that the N/A label (explained as “Data Not Available for This Measure at This Time” in the key or at the bottom of the template) meant that the nursing home in question was deliberately trying to hide poor scores or that it was an indication of

some other problem, such as bad record keeping. Most reported that they would not want to send a loved one to a nursing home with an N/A.

A few respondents questioned whether and how the percentages listed in each table were related to the differences in patient populations among the nursing homes. These respondents recognized that some nursing homes may serve sicker residents and questioned whether the high percentages for these nursing homes would unfairly suggest that they provide worse care. Information on case mix and risk adjustment was not included in this testing.

RECOMMENDATIONS

To display quality measures on Nursing Home Compare, consider using a table such as the “Evaluative Table with Words” or the “Evaluative Table with Stars” rather than a bar graph. In this testing, these tables were both found to be among the templates that elicited the fewest errors, were preferred, and were easiest to use. While other templates performed well in some areas, these two templates stood out as clear winners across all three of the categories in this testing. Moreover, they eliminated the common error of interpreting higher percentages to mean better performance. It is also noteworthy that the “Standard Bar Graph” (based on what is currently displayed on Nursing Home Compare) elicited the most errors and was among those most subject to misinterpretation.

If numeric or graphic formats are used for public reporting, consider framing quality measures in a positive direction. While some respondents found some numeric and graphic templates relatively easy to use and understand, the negative direction of the measures remained confusing. This

suggests that consumers will continue to have difficulty interpreting and using this information correctly.

If measures with a negative direction continue to be used for public reporting, consider displaying the information in an evaluative table to reduce critical errors in interpretation. If it is not feasible to change measures to a positive direction in the near future, this research suggests that using an evaluative table (e.g., “Evaluative Table with Words” or “Evaluative Table with Stars”) would reduce the likelihood of misinterpretation. Numeric tables and bar graphs often led respondents to conclude that the worst performing nursing homes (those with the higher percentages) were the best, notwithstanding the warning label at the top.

Allow users to compare several facilities using an evaluative table on one page and then drill down to bar graphs (or numeric tables) to compare one or two selected facilities. Respondents preferred using the evaluative tables to compare many nursing homes and get an initial idea of overall performance. Using an evaluative table also helped them make fewer errors in interpretation. However, many respondents pointed out that they would like more specific information about the homes they were interested in, including their actual performance on specific measures as compared to the State and national average.

DISCUSSION

In the months since this research was conducted, national commitment to public reporting on health care quality has continued to grow. In addition to its expanded set of measures on nursing homes, CMS currently reports facility-specific measures of quality on home health agencies, hospitals, and dialysis facilities, and new measures are planned. Notwithstanding this

commitment to transparency as a means of promoting quality in health care, engaging consumers effectively in the public dialogue around quality remains a challenge. Quality information may be technically complex. Consumers accustomed to thinking of health care in personal terms may not understand how aggregate measures of performance (or aggregate measures of risk) relate to them. They may not be aware of systematic variations in quality, or they may perceive that they have little choice, in any case. Disturbingly, the research reported here further suggests that even when they are engaged, consumers may erroneously interpret quality information without knowing that they are doing so. Our findings also suggest, however, that thoughtfully designed reporting formats can reduce serious errors and enhance comprehension. Although the recommendations offered here address reporting of quality measures on Nursing Home Compare, in particular, they may also inform other and future public reporting efforts.

This research focused on consumers, because www.Medicare.gov is a consumer-targeted Web site. We recognize, however, that consumers do not make decisions alone, nor should they rely on quality measures alone when making decisions. Additional research focusing on information intermediaries (such as hospital discharge planners, nurses, or physicians) would provide further insight on the design of reporting templates and other informational materials to support health care decisions.

ACKNOWLEDGMENTS

The authors wish to thank Lauren Blatt and Alyson Marano Ward for their contribution to this research. We also wish to thank Rosemary L. Lee for her thoughtful insights and continuing interest and support.

REFERENCES

- Barents Group, KPMG Consulting: *Potential Audiences and Uses of Publicly Reported Quality Data*. Final Report to the Centers for Medicare & Medicaid Services. McLean, VA. November 2001.
- Barents Group, KPMG Consulting: *Consumer Testing of the Out-of-Pocket Costs Page of Medicare Personal Plan Finder*. Final Report to the Centers for Medicare & Medicaid Services. McLean, VA. September 9, 2002.
- Barents Group, KPMG Consulting: *Describing Target Audiences for Facility-Specific Quality Information Provided by Medicare*. Final Report to the Centers for Medicare & Medicaid Services. McLean, VA. January 2003.
- BearingPoint, Inc.: *Findings from Consumer Testing of Decision Tools for Medicare.gov*. Final Report to the Centers for Medicare & Medicaid Services. McLean, VA. 2003a.
- BearingPoint, Inc.: *Consumer Testing of Home Health Compare Web Site Prototype*. Final Report to the Centers for Medicare & Medicaid Services. McLean, VA. 2003b.
- Goldstein, E. and Fyock, J.: Reporting of CAHPS® Quality Information to Medicare Beneficiaries. *Health Services Research* 36(3):477-88, July 2001.
- Hibbard, J.H. and Peters, E.: Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches That Facilitate the Use of Information in Choice. *Annual Review of Public Health* 24:413-33, 2003.
- Hibbard, J.H., Slovic, P.; Peters, E.; et al.: Strategies for Reporting Health Plan Performance Information to Consumers: Evidence From Controlled Studies. *Health Services Research* 37(2):291-313, April 2002.
- Schapira, M.M., Nattinger, A.B., and McHorney, C.A.: Frequency or Probability? A Qualitative Study of Risk Communication Formats Used in Healthcare. *Medical Decision Making* 21(6):459-67, 2001.
- StatCorp LP: *Stat Base Reference Manual, Release 7*. Stata® Press. College Station, TX. 2001.
- Vaiana, M.E and McGlynn, E.A.: What Cognitive Science Tells us About the Design of Reports for Consumers. *Medical Care Research and Review* 59(1):3-35, March 2002.

Reprint Requests: Margaret Gerteis, Ph.D., Mathematica Policy Research, Inc., 955 Massachusetts Avenue, Suite 800, Cambridge, MA 02139 E-mail: mgerteis@mathematica-mpr.com

