Report cards often present conflicting information that requires trade-offs. For example, should a parent choose a plan that performs well on children's care but not so well on adult care or choose a plan that performs just average on both?

Performance Measures

Making Health Care Quality Reports Easier to Use

Judith H. Hibbard, DrPH
Ellen Peters, PhD
Paul Slovic, PhD
Melissa L. Finucane, PhD
Martin Tusler, MS

One of the major dilemmas in improving the quality of medical care is that there are no effective strategies for ensuring that those providing high-quality care are rewarded and those providing low-quality care are penalized. That is, health care practitioners and organizations do not face consequences for providing low-quality care, and there are no effective incentives to motivate improvements in care. Furnishing consumers with comparative performance information so that they can make informed health care choices is one of the few strategies for creating consequences for low or high quality of care performance. The assumption is that if enough consumers made informed choices, high performers would be rewarded with a greater market share and poor performers would be penalized with less market share.

Although there is evidence that consumers want comparative quality information, most studies indicate that there is limited use of the data for decision making. In a review of the evidence to date, Mukamel and Mushlin find only minimal impacts of public reports on consumer choice. Even when reports are sent directly to employees at open enrollment time, only about half attend to the reports or even remember having seen them. Although there are many possible reasons for this lack of response, the complexity of the information and the difficulty of processing and using the amount of information in reports are cited most frequently.

Even consumers who have shown interest in comparative performance information may lack an understanding of data. One fundamental issue is the amount of information presented in reports.

Judith H. Hibbard, DrPH, is Professor of Health Policy, Department of Planning, Public Policy and Management, University of Oregon, Eugene, Oregon. Ellen Peters, PhD, is Research Scientist; Paul Slovic, PhD, is Director; and Melissa L. Finucane, PhD, is Research Scientist, Decision Research, Eugene, Oregon; Martin Tusler, MS, is Research Assistant, Department of Planning, Public Policy and Management, University of Oregon. Copyright © 2001 by the Joint Commission on Accreditation of Healthcare Organizations.

The authors wish to thank the Robert Wood Johnson Foundation for Changes in Health Care Financing and Organization (HCFO) initiative and the Health Care Financing Administration for the funding that supported this work.

Please address requests for reprints to Judith H. Hibbard, DrPH, University of Oregon, 1209 University of Oregon, Eugene, OR 97403-1209; phone 541/346-3364; fax 541/346-2040; e-mail jhibbard@uoregon.edu.
Article-at-a-Glance

Background: Although there is evidence that consumers want comparative quality information, most studies indicate that consumers make limited use of the data in decision making. The reasons for the limited use appear to be the complexity of the information and the difficulty of processing and using the amount of information in reports. The purpose of this investigation was to determine whether there are approaches to reporting comparative information that make it easier for consumers to comprehend the information. Further, the degree to which consumers who have a low level of skill can accurately use that information when it is presented in a format that is easier to use was examined.

Methods: The study used an experimental design to examine how different presentation approaches affect the use of information. Participants were randomly assigned to different conditions and were asked to review information and complete a decision task related to using comparative information and making health plan selections. Two separate convenience samples were used in the study: an elderly Medicare sample \((N = 253)\), and a nonelderly sample \((N = 239)\).

Results: The findings indicate that there are data presentation approaches that help consumers who have lower skills use information more accurately. Some of these presentation strategies (for example, relative stars) improve comprehension among the lower skilled, and other strategies (for example, evaluative labels) appear to aid those in the midrange of comprehension skill.

Conclusions: Using these approaches in reporting would likely increase the use of the comparative information and increase the efficacy of reporting efforts.

Although when asked, consumers frequently say that they want more information, in actual practice they find too much information confusing and overwhelming.\(^{11-13}\)

Jewett and Hibbard\(^4\) identified several concepts that consumers struggle with when reviewing quantitative quality information. Consumers often have difficulty interpreting rates and confuse the concept of rates with similar terms such as ratings or concepts such as fees. Consumers also struggle with interpreting aggregate or comparison data and understanding the implications of a population-based approach to measurement.

Regardless of the format used to present physician-level quality data, it is critical that consumers understand the measures that are used in the reports. Consumers' comprehension of quality information is strongly related to its salience or perceived value.\(^{15}\)

The purpose of the study reported in this article was to determine whether there are approaches to reporting comparative information that make it easier for consumers to understand the information. We suggest that some presentation approaches will help consumers use information more accurately than other approaches to presenting the same data. Further, we examine the degree to which an easier-to-use format can help consumers who have difficulty comprehending comparative information to use the information more accurately.

Difficulty of Using Comparative Performance Information for Choice

Using comparative quality information to make choices requires skills in several areas:

- The ability to correctly interpret data;
- The ability to identify the important factors to integrate into a decision;
- Weighting these factors in ways that match one's individual needs and values;
- Making trade-offs; and
- Bringing all the factors together into a choice.

The latter three skills are quite difficult, for the amount of information in report cards is beyond what humans can effectively process and use. Many report cards list more than 15 performance indicators and may compare as many as 20 plans. Moreover, most consumers have other types of information to factor into their choices, such as plan type, benefits and coverage levels, provider panel considerations, geographic locations, and costs. Trying to integrate several different types of variables into a decision increases the complexity and the difficulty. Bringing all the disparate parts together and not leaving out important variables is a further challenge. The need to make trade-offs adds another level of complexity.\(^{16}\) Within any one option there are likely to be positive and negative elements. Plans do not neatly sort themselves into those that perform well on all indicators and
those that perform poorly on all indicators. This would certainly make the choice easier and clearer. It is more likely that there will be conflicting information that requires trade-offs. As the number of variables for comparison increases, so does the likelihood of the need for trade-offs. For example, should a parent choose a plan that performs well on children's care but not so well on adult care or choose a plan that performs just average on both? Or the trade-off can be across categories of variables (for example, a plan that is low cost but has relatively poor outcomes compared with a high-priced plan with average outcomes). Trade-offs are particularly burdensome and tend to create anxiety and discomfort for the decision maker.

A vast program of research on information processing and human judgment studies demonstrates that when faced with too much information to process or decisions that involve burdensome cognitive processes such as trade-offs, people adopt strategies—shortcuts—to reduce that burden. These shortcuts often take the form of simply choosing based on one factor and ignoring other important factors, thereby possibly undermining the decision maker's self-interest.17

Thus, the complexity and the cognitive difficulty of the tasks involved in using comparative reports is apparently a significant barrier to using the information for making a choice. More recently we have observed that there are significant difficulties for some consumers with the lowest-level skill listed (that is, the ability to correctly interpret comparative performance information). An earlier analysis indicates that more than half of Medicare beneficiaries 65 years of age and older (65+) have significant difficulties correctly interpreting comparative tables and charts showing health plan performance information.18

Which Approaches to Reporting Make Comparative Data Easier to Understand?

In a previous analysis of the data used in this study, we examined the degree to which Medicare beneficiaries could accurately use comparative information to make choices. We later made the same assessments with a nonelderly population. To do these assessments, we developed the comprehension index, which measures the degree to which comparative performance information is understood and used accurately.19 The comprehension index examines the extent to which an individual understands information presented in tables, in charts, and in text form. It assesses the ability to identify optimal choices when viewing unambiguous data.

We found that elderly Medicare beneficiaries make about three times as many errors in interpreting comparative data as do nonelderly consumers (an error was defined as choosing a lower-performing plan within any cost stratum). Further, we found a high degree of variability of skill within this older population. Some nonelderly consumers also have difficulty using comparative information, although there is less variation in skill level within that population.18

The present analysis builds on this earlier work and is informed by the principle of “evaluability,” which asserts that more weight will be given to an attribute in a choice when it is easier to map that information onto a good/bad scale. In other words, making information more valuable means making it easier for people to sort out better and worse options. When information is valuable, the high and low performers pop out at the viewer. Almost no effort is involved in determining what the best options are. For example, if 10 health plans were compared on one measure of performance and if the plans were shown ordered (highest performer to lowest performer), it would be immediately apparent which are the better and worse options in that list.

When information is made more valuable, it is more likely to be used in decision making. An important feature of evaluability is that it operates outside the decision maker's conscious awareness. Unless quality data are made valuable through careful attention to context and formatting, it may fall by the wayside, having far less influence on consumers’ decisions than consumers would like it to have and far less influence than they think it actually did have.1 In a previous study we found that valuable presentation formats, such as ordering by performance and providing visual cues in the form of stars, resulted in greater weighting of quality information in choices.19

In the present study, we examined whether more valuable presentation approaches will also help consumers more accurately understand comparative information in choices. Our specific research questions were as follows:

- Do data presentation approaches designed to be more valuable help consumers use comparative data more accurately?
Example of the Decision Task

<table>
<thead>
<tr>
<th></th>
<th>HMO A</th>
<th>HMO B</th>
<th>HMO C</th>
<th>HMO D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly premium</td>
<td>$50</td>
<td>$75</td>
<td>$48</td>
<td>$63</td>
</tr>
<tr>
<td>Copayment for office visit with primary care doctor</td>
<td>$10</td>
<td>$5</td>
<td>$15</td>
<td>$10</td>
</tr>
</tbody>
</table>

Question 1. Which HMO requires the lowest copayment for a visit with a primary care doctor? (Check one box)
- [ ] HMO A
- [ ] HMO B
- [ ] HMO C
- [ ] HMO D

Figure 1. This figure shows one of the tasks used to create the comprehension index.

Do data presentation approaches designed to be evaluative help consumers who have lower comprehension skill use the information more accurately?

Methods

The data for this study were collected from two separate study populations: a Medicare 65+ group and an employed-age group. The data were used for an earlier analysis that examined the degree to which these population samples were able to correctly interpret comparative data when they were presented in different formats. This analysis uses that same data set to answer the specific research questions listed previously.

Design

The study used an experimental design to examine how different presentation approaches affect the use of health plan attribute information. Participants were randomly assigned to different experimental conditions and were asked to review information and complete a series of decision tasks involving using comparative information and making health plan selections.

Samples

All participants completed a series of tasks designed to test their ability to comprehend unambiguous comparative information. Two separate convenience samples were used in the study: an elderly 65+ Medicare sample (N = 253) and a nonelderly sample (N = 259). Aside from the requirement that participants be at least 65 years of age and be covered by Medicare, there were no other restrictions on participation in the elderly sample. The Medicare beneficiaries were recruited primarily at nonresidential senior centers in Eugene and Springfield, Oregon. Flyers were posted in the senior centers for a week in advance of each session. This portion of the experiment took place between December 1999 and March 2000. The tasks took between 50 and 90 minutes for each respondent to complete.

The nonelderly group was recruited at the University of Oregon from among the nonfaculty staff by mailing a flyer to each eligible person. The only requirements for participation were being younger than 65 years of age and having had health insurance in the past year. The experiment was conducted in July 2000 and took respondents 35 to 70 minutes to finish. All the participants, who were paid, were asked to review the same information as the elderly participants and complete several decision tasks related to using comparative information in making health plan selections. The only difference was that the Medicare group received information printed in a larger font to accommodate age-related declines in vision.

Table 1 (p 596) shows the characteristics of the two study samples and also shows how they compare to a corresponding nationally representative sample.

The Medicare sample. The mean age of participants was 75 years (range, 65–94 years). Sixty-one percent of the study population was women. Twenty-four percent of participants had a college degree or higher; only 9% had less than a high school education. Only 17% of the study population rated their health as fair or poor. This sample was younger, had higher educational levels, and reported better health than the Medicare population as a whole.

The nonelderly sample. The average age of the participants was 39.8 years (range, 18–64 years). The sample was skewed toward women (77%) and persons in better health. The nonelderly sample was also better educated than the same age group in the general U.S. population.

Variables

Comprehension index. The main dependent variable, the comprehension index, assesses the ability to
accurately interpret and use comparative health plan performance information. The comprehension index summarizes performance on 35 decision tasks. These tasks involved interpretation of data presented in different ways, including text, bar graphs, and tables with numbers, and interpretation of positive and negative trends in performance and of data displays that use symbols instead of numbers (see Figures 1–6 for examples of tasks). The score on the comprehension index can range from 0 to 35 and represents the number of errors made in interpreting unambiguous data and/or in making suboptimal choices (such as choosing a lower-performing plan that costs as much as a higher-performing plan). The score on the comprehension index measures the respondent's ability to accurately use comparative information; lower scores indicate greater comprehension (that is, fewer errors made). The index is highly reliable (Cronbach's alpha = .90). Respondents did all the tasks included in the comprehension index. All the data presented in the various tasks in the comprehension index were unambiguous. That is, there were clear right answers or at least one option that had a higher score than the rest.

Figure 7 (p 598) shows how each group is distributed on the comprehension index. Medicare beneficiaries made almost three times as many errors as the nonelderly respondents. The average rate of errors for the nonelderly (on 35 decision tasks) was 9%; for Medicare beneficiaries it was 25%. In addition, performance was more variable in the Medicare population than in the nonelderly population. Among the Medicare sample, those in the quartile with the most errors on the comprehension index averaged 57% errors; those in the quartile with the fewest errors averaged only 5% errors.
## Table 1. Characteristics of the Two Study Samples

| Characteristic | Medicare Sample (N = 253) | MCBS* | Employed-Age Sample (N = 239) | U.S. Census† 
|---------------|---------------------------|-------|-----------------------------|-------------------
| Age           |                           |       |                             |                   
| 18 to 34      | 81%                       | 73%   | 35%                         | 39%§              
| 35 to 64      | 19%                       | 27%   | 65%                         | 61%§              
| 65 to 79      |                           |       |                             |                   
| 80 or more    |                           |       |                             |                   
| Education     |                           |       |                             |                   
| Less than high school | 9%   | 40%   | 3%                          | 11%§              
| High school diploma | 31%  | 32%   | 15%                         | 33%§              
| Some college/vocational school | 37%  | 14%   | 41%                         | 29%§              
| College graduate or more | 24%  | 14%   | 41%                         | 27%§              
| Sex           |                           |       |                             |                   
| Female        | 61%                       | 56%   | 77%                         | 47%§              
| Self-Reported Health |           |       |                             |                   
| Excellent / Very Good | 44%  | 45%   | 61%                         | N/A               
| Good          | 40%                       | 30%   | 32%                         | N/A               
| Fair/Poor     | 17%                       | 25%   | 8%                          | N/A               
| Household Income |                     |       |                             |                   
| Less than $20,000 | 50%  | 60%   | 27%                         | 25%§              
| $20,000 to $39,999 | 34%  | 29%   | 35%                         | 26%§              
| $40,000 to $59,999 | 11%  | 5%    | 22%                         | 19%§              
| $60,000 or more | 5%                        | 6%    | 16%                         | 30%§              
| Marital Status |                             |       |                             |                   
| Married       | 54%                       | 56%   | 64%                         | 60%**             

* MCBS, Medicare Current Beneficiary Survey.  
† U.S. Census Bureau, 1998 Current Population Reports.  
§ Includes only those currently working and aged 18 to 64 years of age.  
ª Includes all households.  
** Includes only those aged 18 to 64 years of age.

Among the nonelderly sample, the quartile with the lowest error rate averaged no errors, and the quartile with the highest error rate averaged 26% errors.  

Quartile scores on the comprehension index and combining the two population scores. As a way of norming the two populations and increasing the n's in the analysis, the two populations were combined for a portion of the analysis. Each participant was assigned a quartile score on the basis of his or her comprehension score. For example, a Medicare consumer who had an error rate of 57% would be in the fourth quartile and receive a quartile score of 4. A nonelderly participant who had a comprehension score of 26% would also be in the fourth quartile for the nonelderly group and would receive a quartile score of 4.

Each study population was divided into quartiles (Figure 7) on the basis of its score on the comprehension index.  

**Predictor variables.** The predictor variables in the analysis are the different data presentation approaches:  
- With (Figure 2) and without (Figure 3) relative stars;  
- In table format without stars (Figure 4, p 597);  
- Ordering the plans by performance (one performance measure; Figure 5, p 598) within premium cost strata versus plans that are alphabetized (unordered by performance or cost; Figure 6, p 598); and  
- With and without labels on bar charts (Figure 8, p 599).
Example of the Decision Task (Table Format Without Stars)

<table>
<thead>
<tr>
<th></th>
<th>Dissatisfied</th>
<th>Neither Satisfied nor Dissatisfied</th>
<th>Somewhat Satisfied</th>
<th>Very or Completely Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan D</td>
<td>18%</td>
<td>8%</td>
<td>25%</td>
<td>49%</td>
</tr>
<tr>
<td>Plan M</td>
<td>16%</td>
<td>7%</td>
<td>26%</td>
<td>51%</td>
</tr>
<tr>
<td>Plan H</td>
<td>13%</td>
<td>5%</td>
<td>20%</td>
<td>61%</td>
</tr>
<tr>
<td>Plan C</td>
<td>20%</td>
<td>6%</td>
<td>31%</td>
<td>43%</td>
</tr>
</tbody>
</table>

Question 1. Given the information above, which plan would you choose? (Check 1 box)

- Plan D
- Plan M
- Plan H
- Plan C

Figure 4. This figure is an example of a task to test the effect of the least evaluable stimulus (table) with the more evaluable tasks (Figures 2 and 3) on decision making.

Example of the Decision Task: Ordering Data by Performance (Ordered Version)

Below are 15 health plans. For each plan you have (a) the monthly cost to be paid by you (above the cost paid by your employer) and (b) the distribution of member ratings on the question "All things considered, how satisfied are you with your current HMO?".

Please examine this information carefully and indicate your first, second, and third preferences by placing a 1, 2, or 3 in the space to the left of the three plans you select.

<table>
<thead>
<tr>
<th>Mark Preferred Plans 1, 2, and 3 Below</th>
<th>Cost</th>
<th>Percentage of Members Who Were:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dissatisfied</td>
</tr>
<tr>
<td>Plan E</td>
<td>$100</td>
<td>7%</td>
</tr>
<tr>
<td>Plan H</td>
<td>$100</td>
<td>13%</td>
</tr>
<tr>
<td>Plan B</td>
<td>$100</td>
<td>10%</td>
</tr>
<tr>
<td>Plan J</td>
<td>$100</td>
<td>12%</td>
</tr>
<tr>
<td>Plan K</td>
<td>$100</td>
<td>14%</td>
</tr>
<tr>
<td>Plan D</td>
<td>$75</td>
<td>14%</td>
</tr>
<tr>
<td>Plan F</td>
<td>$75</td>
<td>14%</td>
</tr>
<tr>
<td>Plan I</td>
<td>$75</td>
<td>13%</td>
</tr>
<tr>
<td>Plan N</td>
<td>$75</td>
<td>13%</td>
</tr>
<tr>
<td>Plan G</td>
<td>$75</td>
<td>17%</td>
</tr>
<tr>
<td>Plan M</td>
<td>$50</td>
<td>16%</td>
</tr>
<tr>
<td>Plan A</td>
<td>$50</td>
<td>15%</td>
</tr>
<tr>
<td>Plan C</td>
<td>$50</td>
<td>20%</td>
</tr>
<tr>
<td>Plan O</td>
<td>$50</td>
<td>20%</td>
</tr>
<tr>
<td>Plan L</td>
<td>$50</td>
<td>22%</td>
</tr>
</tbody>
</table>

Figure 5. This figure is an example of a task to test the effect of a more evaluable stimulus (ordering) with a task with lower evaluable (Figure 6) on decision making.
Example of the Decision Task: Ordering Data by Performance (Unordered Version)

Below are 15 health plans. For each plan you have (a) the monthly cost to be paid by you (above the cost paid by your employer) and (b) the distribution of member ratings on the question "All things considered, how satisfied are you with your current HMO?"

Please examine this information carefully and indicate your first, second, and third preferences by placing a 1, 2, or 3 in the space to the left of the three plans you select.

<table>
<thead>
<tr>
<th>Mark Preferred Plans 1, 2, and 3 Below</th>
<th>Cost</th>
<th>Dissatisfied</th>
<th>Neither Satisfied nor Dissatisfied</th>
<th>Somewhat Satisfied</th>
<th>Very or Completely Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan A</td>
<td>$50</td>
<td>16%</td>
<td>7%</td>
<td>30%</td>
<td>47%</td>
</tr>
<tr>
<td>Plan B</td>
<td>$100</td>
<td>10%</td>
<td>7%</td>
<td>27%</td>
<td>56%</td>
</tr>
<tr>
<td>Plan C</td>
<td>$50</td>
<td>20%</td>
<td>6%</td>
<td>31%</td>
<td>43%</td>
</tr>
<tr>
<td>Plan D</td>
<td>$75</td>
<td>14%</td>
<td>4%</td>
<td>24%</td>
<td>57%</td>
</tr>
<tr>
<td>Plan E</td>
<td>$100</td>
<td>7%</td>
<td>4%</td>
<td>20%</td>
<td>69%</td>
</tr>
<tr>
<td>Plan F</td>
<td>$75</td>
<td>14%</td>
<td>7%</td>
<td>25%</td>
<td>55%</td>
</tr>
<tr>
<td>Plan G</td>
<td>$75</td>
<td>17%</td>
<td>6%</td>
<td>25%</td>
<td>53%</td>
</tr>
<tr>
<td>Plan H</td>
<td>$100</td>
<td>13%</td>
<td>5%</td>
<td>20%</td>
<td>61%</td>
</tr>
<tr>
<td>Plan I</td>
<td>$75</td>
<td>13%</td>
<td>8%</td>
<td>26%</td>
<td>54%</td>
</tr>
<tr>
<td>Plan J</td>
<td>$100</td>
<td>12%</td>
<td>5%</td>
<td>25%</td>
<td>58%</td>
</tr>
<tr>
<td>Plan K</td>
<td>$100</td>
<td>14%</td>
<td>5%</td>
<td>22%</td>
<td>58%</td>
</tr>
<tr>
<td>Plan L</td>
<td>$50</td>
<td>22%</td>
<td>11%</td>
<td>31%</td>
<td>36%</td>
</tr>
<tr>
<td>Plan M</td>
<td>$50</td>
<td>16%</td>
<td>7%</td>
<td>26%</td>
<td>51%</td>
</tr>
<tr>
<td>Plan N</td>
<td>$75</td>
<td>13%</td>
<td>7%</td>
<td>27%</td>
<td>53%</td>
</tr>
<tr>
<td>Plan O</td>
<td>$50</td>
<td>20%</td>
<td>10%</td>
<td>31%</td>
<td>39%</td>
</tr>
</tbody>
</table>

Figure 6. This figure is an example of a task to compare the effect of a less evaluable stimulus (unordered) with a task with higher evaluability (Figure 5) on decision making.

Average Errors by Quartiles of Performance on Comprehension Index

Figure 7. This figure compares the error rates of the two samples when each is divided into error rate quartiles.

Analytic Approach

First, we examined how the different data display approaches affected how well the information was understood and used by the two study populations. We used analysis of variance to assess the data display approach and the degree to which it was related to error rates. We then looked at how much the more evaluable approaches actually helped those with less skill to use the information more accurately. We examined differences in error rates when the data were
more or less evaluable for those at each quartile of performance (on the comprehension index).

Results

First, we examined the degree to which the data display approach affected how accurately the information is interpreted for each of the study samples. Three experiments assessed accuracy in using the data. The fourth experiment was different; we used evaluative labels (excellent, good, fair, poor) to examine the degree to which consumers relied on them in making choices. Thus, the fourth experiment focused on assessing what influences the decision maker rather than on assessing how accurately the information was interpreted.

Then we examined the degree to which more evaluable data presentation approaches help those in the different quartiles of performance (on the comprehension index) use the comparative data more accurately.

Do data presentation approaches designed to be more evaluable help consumers use comparative data more accurately? In the first experiment we examined the impact on accuracy of using relative stars (★★★, better than average; ★★, average; ★, worse than average) in data displays (for example, alongside bar charts) versus just the bar charts without the relative stars (Figures 2 and 3). Relative stars were hypothesized to help the viewer use the information more accurately because the better and worse options are more immediately apparent. The viewer need only identify the options with three stars to see which are the high performers (when the stars are included), making interpretation of the numbers or the bars unnecessary.

In this experiment, four health plans are compared on one measure of member satisfaction. The data were unambiguous, with one plan having the highest score. Participants were asked to identify the best-performing plan. The findings indicate that the relative stars improve the accuracy of using the data among the elderly Medicare sample but not the nonelderly sample. Among the Medicare population, those shown the star version had an average error rate of 18%, whereas those who had no stars in their data
display had an average error rate of 24%. This difference was statistically significant ($p < .05$). For the nonelderly sample, who had considerably fewer errors overall, the relative stars did not increase accuracy in interpreting the data. Nonelderly participants who received the relative star version and those who received the no star version each had error rates of 7%.

In the second experiment, participants were given comparative data shown either in bar charts or in tabular form, without accompanying stars. Four health plans were compared on one dimension of member satisfaction. The data were unambiguous, with one plan having the highest score. We hypothesized that the bar charts would be more evaluable than the numeric tables because the bars’ differing lengths would make it easier to quickly identify the high-performing plans.

However, providing the data in tabular form versus providing the comparative data in bar charts did not have a statistically significant impact on accuracy for either the elderly Medicare sample or the nonelderly sample. The elderly Medicare participants given the table version of the report had a 23% error rate, whereas the Medicare participants given the bar chart version had a 19% error rate in identifying the best performing health plan. Nonelderly participants had an 8% error rate with tables and a 5% error rate with bar charts.

In the third experiment, participants were shown a comparison table of 15 plans. Both monthly premium cost and one performance dimension (member satisfaction) were shown for each health plan (Figures 5 and 6). One group was given the list of plans ordered by performance within cost strata. The other group was given the same list alphabetized (unordered). We hypothesized that ordering plans by performance would make it easier to identify the better and worse options and improve comprehension of the data.

Ordering the plans by performance did significantly decrease errors for the Medicare sample. The error rate was 30% for those with the ordered version and 46% for the unordered version ($p < .01$). The same pattern was observed in the nonelderly sample (15% and 21%), but the differences were not statistically significant.

The fourth experiment did not examine accuracy as the dependent variable but rather the degree to which evaluative labels influenced choices. The question here was whether participants, if given help in the form of evaluative labels, would rely on those labels in making their choices. In this experiment participants were asked to make a choice among two health plans (Figure 8). They were given comparative data on two quality dimensions (consumer ratings of overall satisfaction and the quality of care). One group was given the comparative data in the form of a bar chart; the other group received the same bar chart with added evaluative labels (excellent, good, fair, and poor). We hypothesized that respondents who were relying on the evaluative labels to aid them would more often choose Plan A, as the scores on Plan A cross into the “good” region (when the bar has labels) on both performance measures (Figure 8). The findings indicated that neither the elderly Medicare sample nor the nonelderly sample was significantly influenced by the evaluative labels in their choices.

Do the more evaluable data presentation approaches help consumers who have lower comprehension skills use the information more accurately? In the next step in the analysis we examined whether these same data display approaches help those with lower skills (that is, those who have more errors on the comprehension index) use the data more accurately.

The two populations were divided into quartiles on the basis of their scores on the comprehension index, with the fourth quartile being the group with the most errors. The analyses were carried out for each population separately and then again for a combined sample. Figure 9 (p 601) shows the error rates for each quartile in the combined sample according to whether they were given reports with relative stars. Those in the fourth quartile who were given relative stars had significantly fewer errors in using the report than fourth-quartile participants who were in the no star condition. Second-quartile participants appeared to also be helped by the star format. Scores for those in the first quartile were unaffected by whether they were given the star format. When we examined the elderly Medicare and nonelderly samples separately, a similar pattern emerged, although the effects did not reach statistical significance in either group. Figure 10 (p 601) shows the error rates for each quartile in the combined sample, according to whether they were given reports with tables or with bar charts. Error rates for fourth-quartile participants who viewed the tables were
Figure 9. This figure shows the effect of stars on decision accuracy among comprehension quartiles (combined samples); ns, nonsignificant.

Figure 10. This figure shows decision accuracy (compared with tables) among comprehension quartiles (combined samples); ns, nonsignificant.
higher, although not significantly, than for those who viewed the bar charts. A similar pattern and a finding of no statistical significance were observed for Medicare and the nonelderly when those samples were analyzed separately.

Figure 11 (p 603) shows the error rates for those at each skill level in the combined sample, according to whether they were given reports ordered by plan performance. It appears that ordering helped participants at all skill levels. Within each quartile, those who received the data ordered by performance had lower error scores. These differences were all statistically significant ($p < .05$), with the exception of those in the second quartile. In separate analyses of the elderly and nonelderly samples, the pattern was similar to that shown in Figure 11. However, for the first-quartile nonelderly (that is, the younger high-skill group), ordering did not improve their scores (whether ordered or not, they had few errors).

The last experiment examined the impact of evaluative labels on those at different skill levels (shown in Figure 8). We used evaluative labels (excellent, good, fair, poor) to examine the degree to which consumers rely on these labels in making choices. This experiment focused on assessing what influences the decision maker rather than on assessing how accurately the information is interpreted.

Figure 12 (p 603) shows the percentage of the combined sample that relied on the labels in making choices. Overall the second and third quartiles (mid-to low-level skill groups) seemed to rely relatively more on the labels, but only the third-quartile group showed a statistically significant effect ($p < .05$). When we examined the two samples separately, a somewhat different pattern emerged for each one. Among the nonelderly sample, higher-skill (first- and second-quartile) participants did not rely on the labels for choice (not shown). Lower-skill participants showed some reliance, but this effect did not reach statistical significance. Among Medicare participants, those with the highest skill and the lowest skill (first and fourth quartiles) did not rely on the labels to make choices.

**Discussion**

The findings indicate that there are data presentation approaches that help consumers who have lower skill interpret information more accurately. Some of these presentation strategies improve comprehension among the lower skilled (for example, relative stars), and other strategies appear to aid those in the mid-range of comprehension skill (for example, evaluative labels). Using data display approaches that help those with less skill apparently has no negative impact on those with more skill. Thus, from a policy perspective, using presentation approaches that make the data easy to evaluate would reach a broader base of users than using approaches that are comprehensible only to those with more skill. Because policy approaches work best when they reach the broadest possible audience, pursuing approaches such as the information presentation formats suggested in this analysis is a high priority.

The strength of using an experimental research design to test these questions is that it allows for a high degree of confidence that the observed outcomes are the result of the experimental manipulation. That is, the experimental design allows us to observe how much comprehension is affected when the information remains constant and only the presentation format varies. However, laboratory experiments also have limitations. The information that we asked participants to respond to was contrived and only partially mirrored real-world situations. Participant responses may also have been influenced by the contrived situation. Controlled experiments may not mirror behavior that would occur in natural settings.

Further, the participants in the study were not random samples but convenience samples, which are not generally representative of the larger population from which they were drawn. The nonelderly were more highly educated than the general populations. In addition, the Medicare sample was younger, in better health, and better educated than the larger Medicare 65+ population. That is, they likely reflect a higher-functioning population than the beneficiary population as a whole. Thus, although the study samples included those in the bottom quartile of performance on the comprehension index, it is likely that the participants do not reflect those with the lowest skills in the larger population. Thus, the data display formats that helped those at the low skill end may or may not help those in the lowest skill group not included in our study.

The tasks we asked participants to complete and the data they viewed were actually less complex and less ambiguous than what is currently being disseminated to consumers to aid their choices. Most comparative reports have many more performance measures and health plan
comparisons than what our participants responded to. Further, most comparative reports contain ambiguities that were not present in the tasks used in the study. In public reports it is common for no one health plan option to score well in all categories. This added ambiguity about what is the best choice makes it more complex and burdensome to use the information in making a choice. When consumers are dealing with a more complex information set, the data display approach may become even more important to those with lower comprehension skills. However, because these more complex real-world situations were not tested, it isn’t clear that improved data display will be sufficient to help consumers, particularly lower-skilled consumers.

Thus, even if comprehension difficulties can be minimized by data display strategies, the complexities of
using the information in making a choice (for example, differential weighting of factors, trade-offs) still confront consumers. The use of decision aids in the form of information intermediaries and/or the provision of computer-aided decision support tools will likely still be needed to help consumers with the complexities involved in using comparative data in choice. Computer-aided decision-support tools can be designed to help the user differentially weight factors and make trade-offs consistent with their stated preferences. Such a tool can reduce the cognitive burden of mentally making trade-offs among variables and integrating multiple factors into a choice.

**Summary**

One interpretation of why consumers have failed to use comparative performance information in making choices is that the data are difficult to comprehend. The present findings provide some weight to that argument and to the notion that if comparative data were easier to understand, they would be used more often to inform consumers' choices. Using these approaches in reporting, along with decision aids, would likely increase the use of the comparative information and increase the efficacy of reporting efforts.

**References**