Response Formats and Satisfaction Surveys for Elders

Nicholas G. Castle, PhD,1 and John Engberg, PhD1

Purpose: A factor common to the results of many satisfaction surveys of elders is a lack of response variability. Increasing response variability may be useful if satisfaction surveys of elders are to be productively used in the future. In this paper, we first examine elders’ preferences between five response formats and then examine the response variability of these five commonly used formats. Design and Methods: Satisfaction, demographic, and Short-Form 36 Health Survey data were self-reported by patients in four outpatient surgery centers during 1998 and 1999. We used six different survey instruments randomly given to 3,122 elders. Five instruments varied in response format (5-item Likert format, 5-item satisfaction format, 5-item valuation format, 4-item Chernoff faces, and 10-item visual analogue format). The sixth survey used all five different response sets, and then it asked for the respondents’ preferences among the different response sets. Results: A total of 2,450 questionnaires were examined (response rate of 78.5%). The response format using four Chernoff faces was liked the least, with only 5% of the respondents preferring this format. The 10-item visual analogue format (10VAF) was liked the most, with 39% of the respondents preferring this format. In addition, 10% more elders thought this format was easier to use than the second-place choice (i.e., 32% vs. 22%). The coefficient of variation for the 10VAF was also higher than those in identical domains using the other response formats. This would seem to indicate that the 10VAF is less prone to a ceiling effect than the other response formats. Implications: Our results show that elders have a preference for some response formats, and from the choices we gave them a 10VAF was preferred. The 10VAF also had more response variability then the other formats we tested.

Key Words: Satisfaction, Response formats, Surveys

Several publications document experiences with satisfaction surveys in long-term care settings (e.g., Applebaum, Straker, & Geron, 2000; Cohen-Mansfield, Ejaz, & Werner, 2000; Kleinsorge & Koenig, 1991; Ryden et al., 2000). Many of these provide validated survey instruments, methods to collect data, and advice on using satisfaction results. Most authors also discuss the importance of satisfaction information in long-term care.

Despite these recent works, research on factors that can help improve the information collected from long-term care settings is still needed. Several authors have noted potential difficulties in obtaining survey information from elders. These include problems with response rates (El-Guebaly, Toews, Leckie, & Harper, 1983), cognition (Simmons et al., 1997), acquiescent response bias (Ross, Steward, & Sincere, 1995), and lack of response variability (Pascoe & Attkisson, 1983).

In fact, a factor common to the results of many satisfaction surveys is a lack of response variability—in the case of elders, the use of the upper end of a response format (Pascoe & Attkisson, 1983). On the one hand, a simple dichotomous scale (e.g., yes or no) is most prone to lack of response variability; this scale is also limited in that it gives no indication of the relative intensity of satisfaction or dissatisfaction. On the other hand, the inclusion of more categories (i.e., using a five-category Likert scale) does not appear to resolve the problem. Elders still tend to use the upper end of a response format. Some research with elders also suggests that this may be problematic as more categories can lead to confusion.

Clearly, for satisfaction information from elders to be used more effectively, it would be advantageous to increase the response variability. One way to do this may be to include an easily understood response format. Asking elders which response formats they prefer and which formats they find easiest to use may be a useful first step in increasing response variability. In addition, empirical analyses measuring the response variability between various formats may be

Address correspondence to Nicholas G. Castle, PhD, RAND, 201 North Craig Street, Suite 102, Pittsburgh, PA 15213-1516. E-mail: CASTLE@RAND.org

1RAND, Pittsburgh, PA.
informative for future and current satisfaction initiatives for elders. With these questions in mind, in this paper we first examine elders’ preferences among five response formats, and then we examine the response variability of these five commonly used formats.

Background

Measures of satisfaction have been shown to be useful in many ways. First, results from satisfaction surveys are measures of accountability. Second, satisfaction measures provide customer-focused quality-of-care information that is generally not available from other sources. Third, satisfaction information can be used by providers for quality-improvement initiatives.

The stated goals of organizations promoting and monitoring quality stress the accountability of satisfaction measures. For example, the National Committee on Quality Assurance (NCQA) states that its mission “is to provide information that enables purchasers and consumers of managed health care to distinguish among plans based on quality, thereby allowing them to make more informed health care purchasing decisions” (NCQA, 2003). By publishing satisfaction scores from nursing homes, states such as Michigan, Ohio, and Vermont (Lowe, Lucas, Castle, Robinson, & Crystal, 2003) have also increased the accountability of providers to consumers.

Various structure, process, and outcome measures are addressed in existing facility surveys in long-term care (Cherry, 1991; Harrington, 1991). These are important, especially those measuring clinical outcomes. However, these outcomes may differ from outcomes defined as important by elders (Bond & Thomas, 1992). Moreover, consumers can find clinical outcomes difficult to understand (American Association of Retired Persons [AARP], 2001).

Satisfaction information can be used by providers. The information is useful in at least two ways. First, satisfaction information can be used by providers to compare the quality of care of their own facilities with that of others, although clearly this is most useful when providers are using the same satisfaction instruments. Second, satisfaction information can also be used by providers for quality-improvement initiatives (Harrington, 1991). Noelker, Ejaz, and Schur (2000) found that 98% of nursing homes collecting satisfaction information used the information for quality improvement.

These benefits of satisfaction information rely heavily on adequate and reliable discrimination between satisfaction scores. Clearly, satisfaction surveys with a lack of response variability will provide limited information for accountability, use by consumers, benchmarking, or provider quality-improvement purposes.

Lack of response variability is an issue inherent to health care surveys in general (Pascoe & Attkisson, 1983). Ware, Snyder, and Wright (1977), for example, found approximately 90% of consumers to be satisfied with health care services. Lack of variability would appear to be especially problematic in surveys involving elders, where most elders respond that they are satisfied (Ross et al., 1995). This, of course, is only truly an issue of concern if elders are dissatisfied with the quality of health care.

No gold-standard measure of resident satisfaction exists. However, two factors led us to believe that elders are dissatisfied with the quality of health care, and measurement problems are contributing to lack of response variability in surveys of elders. First, the high satisfaction scores found in most studies conflict with public opinion polls (AARP, 2001), government studies (Institute of Medicine [IOM], 2001), and research (Cherry, 1991) on the general quality of long-term care.

Second, the general survey literature has identified significant associations between the methods of measuring satisfaction and the results obtained. Ware and associates (1977) examined survey-response variability in hospital patients. They used five different measurement approaches for the same participants. The positive responses varied from 1% to 77%, depending on the scale used. In a similar manner, using seven different satisfaction instruments at a Veterans Administration outpatient clinic, Ross and colleagues (1995) found positive responses to vary from 63% to 82%. Counte (1979), examining responses from outpatients at a multiple sclerosis center, found lower satisfaction scores with a 5-point scale than with a 6-point scale. Ware and Hays (1988), using outpatient data from general medical clinics, compared the response variability, reliability, and validity of a 6-point response scale and a 5-point response scale. The 5-point response scale had significantly more variability. These studies would seem to indicate that methods of measuring satisfaction influence the results obtained. However, for the most part, these studies confound different measurement approaches with different response scales, and, as one would expect, given the diverse populations examined, few generalizations can be made from these study findings.

Ross and colleagues (1995) examined the variability in satisfaction evaluations by using seven different measurement methods. These authors conclude that “unreliability of measurement may be a significant problem in satisfaction measurement, especially for the oldest and most ill patients” (p. 392). Despite the conclusion of these authors, no study has examined the association between lack of response variability in surveys of elders and the satisfaction scale used.

The published literature on satisfaction instruments in long-term care we identified are summarized in Table 1. The literature consists of 20 studies, but few generalizations can be made from this table. The
The number of items included in the surveys varies considerably, from as few as 10 (Mitchell & Kemp, 2000) to as many as 79 (Ejaz, Straker, Fox, & Swami, 2003). The mean number of items is 34. Likewise, the number of respondents is highly varied. One study had only 14 respondents (Smith & Sullivan, 1997), whereas the largest number was 1,374 (Pearson, Hocking, Mott, & Riggs, 1993). The mean number of respondents was 232. Six of these publications provided reliability or validity information of their survey instruments. The majority of studies we identified in our literature review were set in nursing homes. Most of these studies are reviewed in the text by Cohen-Mansfield and colleagues (2000), which

Table 1. Characteristics of Long-Term-Care Satisfaction Surveys in the Published Literature

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Name of Instrument</th>
<th>No. of Items in Instrument</th>
<th>No. of Elders in Study</th>
<th>Reliability or Validity Reported</th>
<th>Long-Term Care Setting</th>
<th>Type of Rating Scale</th>
<th>No. of Items in Response Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenwald &amp; Linn (1971)</td>
<td>NHRS</td>
<td>10</td>
<td>26</td>
<td>No</td>
<td>Nursing home</td>
<td>Satisfaction</td>
<td>3</td>
</tr>
<tr>
<td>Pablo (1975)</td>
<td>PSS</td>
<td>68</td>
<td>186</td>
<td>No</td>
<td>Extended care facility</td>
<td>Yes–No</td>
<td>2</td>
</tr>
<tr>
<td>McCaffree &amp; Harkins (1976)</td>
<td>—</td>
<td>56</td>
<td>400</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Satisfaction</td>
<td>3</td>
</tr>
<tr>
<td>Forgan-Morle (1984)</td>
<td>—</td>
<td>29</td>
<td>22</td>
<td>No</td>
<td>Hospital geriatric unit</td>
<td>Yes–No</td>
<td>2</td>
</tr>
<tr>
<td>Cryns et al. (1989)</td>
<td>OPSS</td>
<td>60</td>
<td>229</td>
<td>Yes</td>
<td>Elders in HMOs</td>
<td>Likert</td>
<td>5</td>
</tr>
<tr>
<td>Kurzich et al. (1992)</td>
<td>Satisfaction With Nursing Home Scale (modified)</td>
<td>17</td>
<td>289</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Agree–Disagree</td>
<td>2</td>
</tr>
<tr>
<td>Zinn et al. (1993)</td>
<td>NHRSS Resident Satisfaction Schedule</td>
<td>10</td>
<td>168</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Likert</td>
<td>4</td>
</tr>
<tr>
<td>Norton et al. (1996)</td>
<td>CARE-Q</td>
<td>50</td>
<td>14</td>
<td>No</td>
<td>Nursing home</td>
<td>N/A</td>
<td>—</td>
</tr>
<tr>
<td>Smith &amp; Sullivan (1997)</td>
<td>NHSQI</td>
<td>32</td>
<td>103 and 194</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Likert</td>
<td>4</td>
</tr>
<tr>
<td>Davis et al. (1997)</td>
<td>RSI</td>
<td>27</td>
<td>156</td>
<td>Yes</td>
<td>Assisted living</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Sikorska (1999)</td>
<td>—</td>
<td>14</td>
<td>99</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Yes–No</td>
<td>2</td>
</tr>
<tr>
<td>Simmons &amp; Schnelle (1999)</td>
<td>QUALCARE Scale</td>
<td>54</td>
<td>301</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Evaluation</td>
<td>5</td>
</tr>
<tr>
<td>Bravo et al. (1999)</td>
<td>—</td>
<td>29</td>
<td>No</td>
<td>Long-term care residents</td>
<td>Yes–No</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Fowler et al. (1999)</td>
<td>RSI</td>
<td>42</td>
<td>257</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Yes–No</td>
<td>2</td>
</tr>
<tr>
<td>Uman et al. (2000)</td>
<td>FSQ</td>
<td>10</td>
<td>201</td>
<td>Yes</td>
<td>Assisted living</td>
<td>Satisfaction</td>
<td>7</td>
</tr>
<tr>
<td>Mitchell &amp; Kemp (2000)</td>
<td>—</td>
<td>15</td>
<td>133</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Likert</td>
<td>3</td>
</tr>
<tr>
<td>Ejaz et al. (2002)</td>
<td>—</td>
<td>49</td>
<td>201</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Yes–No</td>
<td>2</td>
</tr>
<tr>
<td>Straker &amp; Ejaz (2001)</td>
<td>—</td>
<td>62</td>
<td>239</td>
<td>Yes</td>
<td>Nursing home</td>
<td>Likert</td>
<td>5</td>
</tr>
</tbody>
</table>

Notes: NHRS = Nursing Home Rating Scale; PSS = Patient Satisfaction Scale; OPSS = Older PSS; NHRSS = Nursing Home Resident Satisfaction Scale; CARE-Q = Caring Assessment Instrument; NHSQI = Nursing Home Service Quality Inventory; RSI = Resident Satisfaction Index; NA = not applicable; FSQ = Facility Satisfaction Questionnaire.
includes a chapter summarizing published satisfaction scales (Kruzich & Cohen-Mansfield, 2000). Of most significance to this paper, we could not find any consensus on the type of response format. Several studies (seven) used a Likert format; however, the number of possible responses in these scales varied from three to five. Other formats commonly used include simple dichotomies, such as yes–no and open-ended response formats.

Methods
Response Formats

The fact that we could not find any consensus on the type of response format used in surveys with elders provided the motivation for this study. Therefore, we undertook an examination of elders’ preferences between different response formats and the response variability of commonly used scales. Our first task was to choose response formats to examine. To do this we examined health care satisfaction surveys for examples of possible response formats (such as the long-term-care surveys listed in Table 1). This review included the rather ubiquitous hospital and managed care surveys commonly used in today’s health care environment. We also included a review of surveys most often used by mental health care providers. We included mental health care providers because of the known cognitive and mental limitations from which many elders suffer.

Some authors have described types of response formats used in satisfaction surveys. For example, Krowinski and Steiber (1996) gave examples of four general categories of response formats: nominal, ordinal, interval, and ratio. We first tried using these conventional categories to classify the response formats of commonly used surveys. We found most response formats to fall into the ordinal category. Thus, because of the lack of discrimination between response formats using this approach, at least for our purposes, we more finely divided satisfaction surveys by using the type of rating scale used (also called scaled-response formats; Krowinski & Steiber, 1996).

We identified eight commonly used scaled-response formats: open ended, dichotomous (e.g., yes–no), Likert, evaluation, frequency, satisfaction, visual analogue, and Chernoff faces. The format of open-ended and dichotomous response formats is self-evident. Likert formats are a series of opinions (e.g., agree, strongly agree; see Likert, 1932). An evaluation format is a series of responses of ordinal value (e.g., poor, good, very good). A frequency format is a series of responses with ascending or descending value (e.g., all of the time, some of the time). A satisfaction format is a scale that determines the frequency of satisfaction (e.g., very satisfied, satisfied). A visual analogue format (also called graphic scaling) is a pictorial scale that usually has some interval value (e.g., scale from 1 to 10). Chernoff faces are a pictorial representation with Likert-scale or evaluation-scale types of values (faces have smiles and frowns). Examples of these eight types of scaled-response formats are given in the appendix.

Because it was not possible to include measures using all potential variations in response scales and response formats, we used five common types. Our criteria for choosing these were as follows: (a) they were commonly used in existing long-term-care surveys; (b) they were commonly used in health care surveys; (c) they were commonly used by mental health care providers; and (d) positive opinions of the scales or formats were published. We excluded open-ended questions because they can be problematic in creating empirical scores. We excluded dichotomous scales because they limit the degree of information obtained on satisfaction–dissatisfaction; for example, the number of satisfied residents can be determined, but no information is gained on how satisfied they are. We also excluded frequency scales because they were inappropriate for many questions we used to evaluate the one-time medical encounter (surgery center visit) investigated. Using this process, we identified five response formats.

First, we used a 5-point Likert format (5LF), anchored by 5 (strongly agree) and 1 (strongly disagree). An example of this was the following request: “I am satisfied with my overall experience in the surgery center” (strongly disagree, disagree, not sure, agree, strongly agree).

Second, we used a 5-point satisfaction format (5SF), anchored by 5 (very satisfied) and 1 (very dissatisfied). An example of this was the following request: “Please rate your overall satisfaction with our surgery center” (very dissatisfied, dissatisfied, neither satisfied or dissatisfied, satisfied, very satisfied).

Third, we used a 5-point evaluation format (5EF), anchored by 5 (excellent) and 1 (poor). An example of this was this request: “Please rate your overall experience in our surgery center” (poor, fair, good, very good, excellent).

Fourth, we used the four Chernoff face format (4CF), anchored by 4 (excellent) and 1 (poor). Under the faces the ratings excellent, very good, fair, and poor were printed. An example of this was the following request: “Please rate your overall experience in our surgery center” (poor, fair, very good, excellent).

Fifth, we used a visual analogue format (10VAF), anchored by 1 (very poor) to 10 (excellent). These scales themselves have a lot of variability. They vary in the anchors used and whether demarcations are included on the scale (Pascoe & Attkisson, 1983; Ross et al., 1995). We included demarcations from 1 to 10, and respondents were asked to mark on the scale the point that best identified their experience.

Data Sources and Sample Selection

Satisfaction and demographic data were self-reported by patients in four outpatient surgery...
centers during 1998 and 1999. In addition, self-reported questions of general health taken from the Short-Form 36 Health Survey (SF-36; Ware & Sherbourne, 1992) were included. The SF-36 has five questions that are combined into a general health perceptions scale. In this investigation only data from patients 65 years or older are reported.

We used six different survey instruments. Five instruments varied in response format, as already described (5LF, 5SF, 5EF, 4CF, 10VAF). In these five satisfaction questionnaires, we kept other factors, such as size of the font and the order of question, constant. The wording of questions was also extremely similar in all cases. However, we could not use identical wording for the questions, as minor changes had to be made to make the question appropriately match the response scale. In the sixth survey instrument, we used questions from two satisfaction domains (art of care and global satisfaction) and asked each question multiple times by using the five different response sets. Then a series of questions asked for the respondents' preferences between the different response sets. We used the art of care and global satisfaction domains, because in pilot analyses these domains had the least missing data.

Surveys were given to participants from a randomly ordered stack. We randomized the order of the surveys by using a table of random numbers. All survey instruments had a face page, so the type of survey given to each patient could not be readily identified. We performed regular checks to ensure that the administration of the different survey instruments was random. In addition, we have no reason to believe that administration of the different survey instruments was biased because of staff preferences.

**Instrument Development and Administration**

In the published literature, we were unable to find a short, valid, satisfaction instrument for use in an outpatient surgery center. Therefore, we developed our own survey by using five criteria from the Centers for Medicare and Medicaid Services (CMS) for assessing the validity of our instrument (construct validity, face validity, reliability, clinical validation, and applicability).

No attempt was made to assess satisfaction with all aspects of outpatient surgery center care. The main criterion was to simply collect reliable and valid information that could be used for quality-improvement purposes. However, in an emulation of previous studies, three dimensions of elder satisfaction were purposefully evaluated: art of care, technical quality, and efficacy. This was based on the work of Ware, Davis-Avery, and Stewart (1978).

The art of care assesses provider characteristics. In our case this includes courtesy and comfort in asking questions. The technical quality of care assesses competence of caregivers. We include the knowledge and explanations given by physicians and nurses. Efficacy assesses the degree to which patients feel they were helped by the care given. We include the care given by physicians and nurses. In addition, we included four questions regarding the amenities of the care environment and one global item. The amenities items included temperature, cleanliness, and comfort of the facility. The global item assessed overall satisfaction with the outpatient surgery center.

We report on the reliability of the instrument in the results section. To ensure face validity we used a multidisciplinary hospital–surgical center team including clinical and administrative staff to develop the survey. In addition, nurses in daily contact with patients were key participants in the project. From a comprehensive list of questions compiled by our team and from a literature search, we discussed which information we actually required. This, we believe, ensured the applicability of the survey.

We sought to use simple language and short questions. In all cases, a Flesh–Kincaid Grade reading level of 5.0 or lower was used (the average for all questions was 4.7). The resulting 17 questions were pilot tested with patients for one month ($n = 112$). This resulted in no changes to the questions and only minor wording revisions to the survey instructions given to the patient. The data from the pilot test are not included in the results we report.

**Survey Administration**

To ensure responses were representative of all the time spent in the outpatient surgery center, we arranged for the survey to be given to patients after recovery from surgery. The potential disadvantage of this approach is that patients may be recovering from the effects of anesthesia and may have some pain. The effects of anesthesia were not problematic in our study. As a normal part of the outpatient surgical experience, patients spend a considerable amount of time in the recovery area so that the effects of anesthesia dissipate, and patients are given pain medication.

Patients were provided with instructions to deposit the completed questionnaire in a locked box in the waiting area, or mail it back with an attached postage-paid envelope. The instructions also described how the information would be used and that completing the survey was voluntary. No signed informed consent was used; however, patients were assured that no individuals would be identified in any use of the data. In fact, it is important to note that no identifying marks were used on the survey, and once it was deposited in the locked box the patient could not be identified.

A potential bias to satisfaction surveys involves recall bias. That is, patients may have difficulty in
accurately responding to some questions. Our survey administration process ensured that patients received a questionnaire when their experience was still memorable and thus decreased bias caused by potential time lags. Other biases may still exist. For example, patients may have a social desirability tendency. That is, patients may respond to the satisfaction questions as they think they are expected to respond. Patients may also have a fear of reprisal, although the anonymity of patients was an essential part of our survey process.

**Analyses**

We first provide some of the psychometric properties of the satisfaction instrument. This follows the work of McHorney, Ware, Lu, and Sherbourne (1994) and includes completeness of data, item-discriminant validity, and reliability of scale scores.

We determined the percentage of patients not providing responses for each question. This is important because a score for each scale cannot be confidently computed if a high number of individual items comprising that scale are missing (McHorney et al., 1994). Item-discriminant validity determines the degree to which items correlate within the scale compared with other scales (see McHorney et al. for a discussion of this technique). Finally, to examine the internal consistency, we calculated Cronbach's alpha for each scale.

Respondents' preferences for each response scale are presented. We use *t* tests, with Tukey corrections to account for multiple comparisons, to compare the significance of the difference in values between the groups.

Using a methodology similar to that of Ware and Hays (1988), we transformed the satisfaction scores from the 5LF, 5SF, 5EF, 4CF, and 10VAF formats to a common 0–100 scale, with higher scores reflecting greater satisfaction. For each scale, item responses in the lowest category were recoded as 0, and those in the highest category were recoded as 100. Item responses between the highest and lowest categories were recoded to give uniform divisions between 0 and 100. For example, using the 5EF, we recoded responses of poor, fair, good, very good, and excellent as 0, 25, 50, 75, and 100, respectively. We also calculated the coefficient of variation (CV) by dividing the standard deviation by the mean and multiplying by 100 (Ware & Hays, 1988). The CV is a unitless measure, and it enables us to compare the response variability of the different satisfaction formats used. There is no simple parametric distribution for a test statistic that would allow us to test the difference between two CV values (Rao & Bhatt, 1995). Therefore, we bootstrapped a 95% confidence interval for each of the CV estimates to determine whether the differences between the CVs are due to more that just sampling error (Davison & Hinkley, 1997).

**Results**

The percent of residents not providing responses for each question was low and varied from 1.5% (overall satisfaction) to 5.1% (care given by doctors helped me get well). The item-discriminant validity analyses determined that the correlation of items within scales was higher than those with other scales. McHorney and colleagues (1994) used a measure of correlation between questions of greater than 2 standard errors in item-discriminant validity analyses. In all but one case, each of our items achieved this level. Cronbach's alphas were 0.79 for the six-item art-of-care scale, 0.72 for the four-item technical quality scale, 0.90 for the two-item efficacy scale, and 0.89 for the three-item amenities scale. A level of 0.70 or greater is usually recommended (McHorney et al., 1994).

A total of 2,450 valid questionnaires were received. Census records indicate that 3,122 elders received care in the outpatient surgery centers during the period of observation. This gives a response rate of 78.5%. In addition, response rates for the five different questionnaires were similar, varying from 76.3% (5SF) to 79.8% (10VAF). Loss of 672 elders may be representative of those with lower levels of satisfaction or more severe illness. As already discussed, some patients may have fear of reprisal in answering the questionnaire, and presumably they are dissatisfied. Other patients may not have responded as a result of their illness. Only 7% of the sample was from mailed questionnaires. In sensitivity analyses excluding these questionnaires, findings were similar to those presented. That is, values excluding these questionnaires were all within 1% of the values reported. In addition, response rates, use of the mail to return the questionnaire, levels of perceived physical health, and average age were not significantly different among the six groups of respondents that received different questionnaire formats (analysis not shown).

Descriptive statistics (not shown) show that, on average, patients rated their overall health as above average (*M* = 4.61). Only 11% of respondents rated their health as very poor. This would be expected, as patients in very poor health are more likely to receive surgery in a hospital setting. The average age of residents was 69 years.

Table 2 presents the results for patient preferences for the survey response formats used. This information came from the sixth survey instrument, which used questions from two satisfaction domains (art of care and global satisfaction) and asked each question multiple times with the five different response sets. The 4CF was liked the least, with only 5% of the respondents preferring this format. The 10VAF was
liked the most, with 39% of the respondents preferring this format.

Table 3 presents the response variability for the five different questionnaire formats used. For all five scales, the 10VAF had the highest CV. In 19 out of 20 possible CV comparisons with other formats (i.e., five scales multiplied by four other formats), the entire 95% confidence interval for the 10VAF CVs is above the confidence intervals for the other scales. This implies that the difference in the CVs is statistically significant. At the other end of the spectrum, the 4CF has the lowest or is tied for the lowest CV for all five scales. However, in most cases the confidence intervals overlap with those of the three 5-point scales. Only for the efficacy scale does the 4CF have significantly less variation than all three of the 5-point scales.

Discussion

Our objective in this paper was first to examine elders’ preferences among five response formats and second to examine the response variability of these five commonly used formats. Our results clearly show that elders have a preference for the 10VAF. Seventeen percent more elders preferred this format than the second-place choice (i.e., 39% vs. 22%). In addition, 10% more elders thought this format was easier to use than the second-place choice (i.e., 32% vs. 22%).

The CV for the satisfaction domains used in the 10VAF was also higher than those in identical domains using the other response formats. This would seem to indicate that the 10VAF is less prone to a ceiling effect than the other formats. In fact, the mean scores for the 10VAF were all approximately 5 points lower than those of the next best format, and the CV was about 3 points higher. We are careful to stress, however, that although the 10VAF would appear to be less prone to a ceiling effect than the other scales, this does not necessarily mean the 10VAF is immune from ceiling effects. It is also important to note that in all cases our results were skewed to the positive end of all response formats.

Our results should also be tempered by several limitations inherent to this investigation. First, we only have data from relatively well-functioning community-dwelling elders. Although we have little data to the contrary, one could envisage elders with low cognitive functioning to be confused with a 10-item format and the extra wording needed to introduce the format (the word stem—on a scale of 1 to 10, please rate ...). Indeed, in nursing home residents with low cognitive functioning, dichotomous scales appear advantageous and are advocated by several researchers (e.g., Simmons et al., 1997; Uman et al., 2000). In addition, in most other long-term care settings, residents are older and frailer than those we investigated. Our results may have limited generalizability and may not be applicable to the institutionalized elderly population.

Table 2. Descriptive Statistics of Patient Preferences for Survey Response Formats

<table>
<thead>
<tr>
<th>Question</th>
<th>5LF (%)</th>
<th>5SF (%)</th>
<th>5EF (%)</th>
<th>4CF (%)</th>
<th>10VAF (%)</th>
<th>Tukey (p &gt; .05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which question did you find easiest to answer?</td>
<td>21</td>
<td>17</td>
<td>18</td>
<td>15</td>
<td>29</td>
<td>10VAF &gt; 5LF</td>
</tr>
<tr>
<td>Which question scale did you find easiest to answer?</td>
<td>22</td>
<td>16</td>
<td>17</td>
<td>14</td>
<td>32</td>
<td>10VAF &gt; 5LF, 5EF, 5SF, 4CF</td>
</tr>
<tr>
<td>If you had to choose one, which question scale do you prefer?</td>
<td>22</td>
<td>17</td>
<td>19</td>
<td>5</td>
<td>39</td>
<td>10VAF &gt; 5LF, 5EF, 5SF, 4CF</td>
</tr>
</tbody>
</table>

Notes: For this table, n = 459. 5LF = 5-point Likert format: strongly disagree, agree, not sure, agree, strongly agree; 5SF = 5-point satisfaction format: very dissatisfied, dissatisfied, neither satisfied nor dissatisfied, satisfied, very satisfied; 5EF = 5-point evaluation format: poor, fair, good, very good, excellent; 4CF = four Chernoff faces: poor, fair, good, excellent; 10VAF = 10-point visual analogue format 1, very poor, and 10, excellent.

The second limitation of our research is that we only examine five different response formats. Our results are probably only representative of these specific formats. We cannot extrapolate from our results, for example, to state that the 10VAF would perform better than a three-item Likert format (not investigated). Nor can we presume that the 10VAF is the best visual analog format. A similar format with 0 to 10 responses (rather than 1 to 10), a format with fewer responses such as from 1 to 5, or a format using different anchors may perform better. In addition, because four of the formats examined used 5 or fewer response options, and only one used 10 response options, it would seem especially appropriate to examine response options in between (e.g., a seven-item response option). In short, the response formats we examine vary on two dimensions: scale format and number of response options. With the number of response formats we tested, we cannot determine whether the 10VAF was favored because it used 10 response options or because it used a visual analogue format. Of course, we also do not investigate why participants preferred the 10VAF.

We examine common satisfaction questionnaire response formats. These response formats have face validity, but we have no way of knowing whether less commonly used response formats would perform better than the 10VAF. We also do not examine response formats that use a dual-rating system, that is, a format using a yes–no question first, followed by a specific response format (such as a Likert scale) if yes is given as an answer.

A further limitation of our research is that some
bias may have come from patients’ relatives, who may have helped in completing the questionnaire. We did not include a section on the survey to identify whether such help was used. In retrospect this may have been useful information to collect.

Several initiatives are currently underway to standardize the collection of satisfaction information in long-term-care settings. The most notable are the nursing home Consumer Assessment of Health Plans Survey (CAHPS; Agency for Healthcare Research and Quality, 2001) and the Ohio Department of Aging instruments (Straker, Ehrichs, Ejaz, & Fox, 2003; Straker & Ejaz, 2001). The full benefits of measuring and reporting resident satisfaction can only be realized with such standardized instruments. Benchmarking, accountability, and quality-of-care information especially require valid and consistent measures. The architects of these satisfaction instruments have a considerable number of pertinent issues to consider in order to ensure the validity of their instruments. These issues include, but are not limited to, which domains to measure, number of questions, length of questions, survey administration, and cognitive screening. We propose that one further consideration to be included in the development of future satisfaction surveys is the response format used.

**Table 3. Response Variability and Satisfaction Formats**

<table>
<thead>
<tr>
<th>Satisfaction Domains</th>
<th>Response Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5LF(1)</td>
</tr>
<tr>
<td>Art of Care</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>82.2</td>
</tr>
<tr>
<td>SD</td>
<td>18.5</td>
</tr>
<tr>
<td>CV (95% CI)</td>
<td>22.5 (21.1–24.3)</td>
</tr>
<tr>
<td>Technical Quality</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>84.4</td>
</tr>
<tr>
<td>SD</td>
<td>18</td>
</tr>
<tr>
<td>CV (95% CI)</td>
<td>21.4 (19.4–23.0)</td>
</tr>
<tr>
<td>Efficacy</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>73.0</td>
</tr>
<tr>
<td>SD</td>
<td>20.3</td>
</tr>
<tr>
<td>CV (95% CI)</td>
<td>27.8 (25.9–30.1)</td>
</tr>
<tr>
<td>Amenities</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>76.3</td>
</tr>
<tr>
<td>SD</td>
<td>20.9</td>
</tr>
<tr>
<td>CV (95% CI)</td>
<td>27.4 (24.7–30.3)</td>
</tr>
<tr>
<td>Global Satisfaction</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>84.5</td>
</tr>
<tr>
<td>SD</td>
<td>20.3</td>
</tr>
<tr>
<td>CV (95% CI)</td>
<td>24.0 (20.9–27.1)</td>
</tr>
</tbody>
</table>

Notes: All satisfaction scores were transformed to a common 0–100 scale, and the coefficient of variation (CV) was calculated by dividing the standard deviation by the mean (X) and multiplying by 100. The confidence interval for the CV is a bias-corrected bootstrap confidence interval (CI) based on 1,000 replications. 1 = 5-point Likert format (5LF): strongly disagree, agree, not sure, agree, strongly agree; 2 = 5-point satisfaction format (5SF); 3 = 5-point evaluation format (5EF); 4 = four Chernoff (4CF) faces; 5 = 10-point visual analogue format (10VAF); 1 was anchored by very poor and the 10 was anchored by excellent; H = highest CV for each satisfaction domain; L = lowest CV for each satisfaction domain. For 5LF, n = 412; 5SF, n = 351; 5EF, n = 388; 4CF, n = 411; 10VAF, n = 429.

**References**


Ejaz, F. K., Straker, J. K., Fox, K., & Swami, S. (2003). Developing...


Ware, J. E., Snyder, M. K., & Wright, W. R. (1977). Some issues in the measurement of patient satisfaction with health care. Santa Monica, CA: RAND.


**Received December 3, 2002**

**Accepted September 23, 2003**

**Decision Editor: Linda S. Noelker, PhD**

---

**Appendix**

**Survey Response Formats**

We classified rating formats as Likert, evaluation, frequency, satisfaction, visual analogue, Chernoff faces, open-ended, and dichotomous. In open-ended questions, “respondents are asked structured questions but are allowed to answer in their own words” (Krowinski & Steiber, 1996, p.132). In dichotomous questions, respondents are given two response options—usually “yes” or “no.”

Likert formats are a series of opinions with equal intervals (often called equal-appearing interval scales), such as the 5-point Likert format (5LF) we used:

- ![Strongly Disagree](Image)
- ![Disagree](Image)
- ![Not Sure](Image)
- ![Agree](Image)
- ![Strongly Agree](Image)

An evaluation format is a series of responses of ordinal value, such as the 5-point evaluation format (5EF) we used:

- ![Poor](Image)
- ![Fair](Image)
- ![Good](Image)
- ![Very Good](Image)
- ![Excellent](Image)
A frequency scale is a series of responses with ascending value:

- All of the time
- Some of the time
- At no time

A satisfaction format is a series of responses that measures the frequency of satisfaction, such as the 5-point satisfaction format (5SF) we used:

- Very Dissatisfied
- Dissatisfied
- Neither Satisfied or Dissatisfied
- Satisfied
- Very Satisfied

A visual analogue scale is a pictorial scale that usually has some interval value, such as the 10 point visual analogue rating format (10VAF) we used:

- Very Poor
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 Excellent

Chernoff faces are a pictorial representation with Likert scale–evaluation scale types of values, such as the four Chernoff faces (4CF) we used:

- Poor
- Fair
- Very Good
- Excellent

“Indispensable to gerontology research” — Harry R. Moody

AgeLine Database
New books and articles in social gerontology

http://research.aarp.org/ageline
(also on Ovid, SilverPlatter, CSA, EbscoHost, Dialog, and NISC)

Other AARP resources: (link from AgeLine site)
- AgeSource Worldwide
  250 global aging resources from 24 countries
- AARP Research Center
  Complete text of publicly available AARP research
- Internet Resources on Aging
  700 web sites for or about older adults—search or browse
- AgeLine Research to Go
  New publications on high-interest topics—great for students!