

A Demonstration of the Impact of Response Bias on the Results of Patient Satisfaction Surveys

Kathleen M. Mazor, Brian E. Clauser, Terry Field, Robert A. Yood, and Jerry H. Gurwitz

Objectives. The purposes of the present study were to examine patient satisfaction survey data for evidence of response bias, and to demonstrate, using simulated data, how response bias may impact interpretation of results.

Data Sources. Patient satisfaction ratings of primary care providers (family practitioners and general internists) practicing in the context of a group-model health maintenance organization and simulated data generated to be comparable to the actual data.

Study Design. Correlational analysis of actual patient satisfaction data, followed by a simulation study where response bias was modeled, with comparison of results from biased and unbiased samples.

Principal Findings. A positive correlation was found between mean patient satisfaction rating and response rate in the actual patient satisfaction data. Simulation results suggest response bias could lead to overestimation of patient satisfaction overall, with this effect greatest for physicians with the lowest satisfaction scores.

Conclusions. Findings suggest that response bias may significantly impact the results of patient satisfaction surveys, leading to overestimation of the level of satisfaction in the patient population overall. Estimates of satisfaction may be most inflated for providers with the least satisfied patients, thereby threatening the validity of provider-level comparisons.

Key Words. Patient satisfaction, response bias, response rate

In recent years, health care organizations, policymakers, advocacy groups, and individual consumers have become increasingly concerned about the quality of health care. One result of this concern is the widespread use of patient satisfaction measures as indicators of health care quality (Carlson et al. 2000; Ford, Bach, and Fottler 1997; Rosenthal and Shannon 1997; Young, Meterko, and Desai 2000). In some organizations, patient satisfaction survey results are used in determining provider compensation (Gold et al. 1995).

As in any measurement procedure, biased results pose a severe threat to validity. Random selection is often used to ensure that patients who *receive* a

questionnaire are representative, but random selection does not ensure that those who *respond* are also representative. Researchers have become increasingly aware that systematic differences between respondents and nonrespondents are a greater cause for concern than low response rates alone (Asch, Jedrzejewski, and Christakis 1997; Krosnick 1999; Williams and Macdonald 1986).

Numerous studies have assessed the differences between responders and nonresponders (or initial responders and initial nonresponders) on demographic variables, respondent characteristics, health status, and health-related behaviors; most have found differences on at least some variables (van den Akker et al. 1998; Band et al. 1999; Benfante et al. 1989; Diehr et al. 1992; Etter and Perneger 1997; Goodfellow et al. 1988; Heilbrun et al. 1991; Hill et al. 1997; Hoeymans et al. 1998; Jay et al. 1993; Lasek et al. 1997; Launer, Wind, and Deeg 1994; Livingston et al. 1997; Macera et al. 1990; Norton et al. 1994; O'Neill, Marsden, and Silman 1995; Panser et al. 1994; Prendergast, Beal, and Williams 1993; Rockwood et al. 1989; Smith and Nutbeam 1990; Templeton et al. 1997; Tennant and Badley 1991; Vestbo and Rasmussen 1992). However, differences between responders and nonresponders are difficult to identify when the difference is on a variable for which prior information is not available for the full sample. For instance, it is relatively straightforward to assess whether older patients respond at a higher rate than younger patients by comparing age distributions of the survey sample and the respondent sample. In contrast, for a variable such as satisfaction, the underlying distribution in the full sample is not known, and so there is no straightforward means of determining whether the distribution of this variable for respondents differs from the distribution for the full sample. However, in spite of the difficulties

Address correspondence to Kathleen M. Mazor, Ed.D., Meyers Primary Care Institute, Fallon Healthcare System and University of Massachusetts Medical School, 630 Plantation Street, Worcester, MA 01605. Dr. Mazor is Senior Research Associate, Meyers Primary Care Institute, Fallon Healthcare System and University of Massachusetts Medical School, and Assistant Professor of Medicine, University of Massachusetts Medical School, Worcester, MA. Brian E. Clauser, Ed.D. is Senior Psychometrician, National Board of Medical Examiners, Philadelphia, PA. Terry Field, D.Sc., is Associate Director, Meyers Primary Care Institute, Fallon Healthcare System and University of Massachusetts Medical School, and Assistant Professor of Medicine at the University of Massachusetts Medical School, Worcester. Robert A. Yood, M.D., is Medical Director for Research at Fallon Clinic Inc., and Associate Professor of Medicine at the University of Massachusetts Medical School, Worcester. Jerry H. Gurwitz, M.D. is Executive Director of the Meyers Primary Care Institute, Fallon Healthcare System and University of Massachusetts Medical School, and Professor of Medicine at the University of Massachusetts Medical School, Worcester.

inherent in this sort of investigation, there are a number of studies that have attempted to assess differences in satisfaction between respondents and nonrespondents (or early and late respondents). Findings from these studies suggest that nonrespondents or late respondents may evaluate care differently, and perhaps less favorably, than respondents or early respondents, respectively (Barkley and Furse 1996; Etter, Perneger, and Rougemont 1996; Lasek et al. 1997; Pearson and Maier 1995; Woolliscroft et al. 1994).

The first objective of the present study was to examine actual patient satisfaction data for evidence of response bias. The second objective was to examine how response bias (in this case, differential likelihood of response as a function of level of satisfaction) might impact patient satisfaction survey results using simulated data.

METHODS

The setting for the patient satisfaction survey was a group-model health maintenance organization, in central and eastern Massachusetts, with a total enrollment of more than 150,000 members. Members are covered for most outpatient medical costs with a nominal copayment for outpatient services and are served by a large multispecialty group practice of more than two hundred physicians.

The present study was limited to family practitioners and general internists. Patient satisfaction questionnaires were mailed to a random sample of patients who visited their primary care provider during the previous three months. For each provider, up to 150 patients per quarter were selected to receive a questionnaire. Patients who had been sent a survey during the previous two years were not eligible. While patient satisfaction with all providers in the medical group is periodically assessed, the organization places a special focus on satisfaction with primary care providers because each health plan member is assigned to a specific primary care provider to manage their health care needs.

Data collection occurred during the spring of 1997 and the spring of 1998. All responses were anonymous; administration consisted of a single mailing, without follow-up. Data from these two periods (chosen to minimize the impact of secular trends) were combined after determining that the average difference in mean satisfaction rating (for the same provider for the two quarters) was approximately .08, and the corresponding correlation (for pairs of satisfaction ratings for the two time periods) was .91. In addition, the

average difference in response rate for providers was 8 percent, and the corresponding correlation (for matched pairs of response rates) was .58.

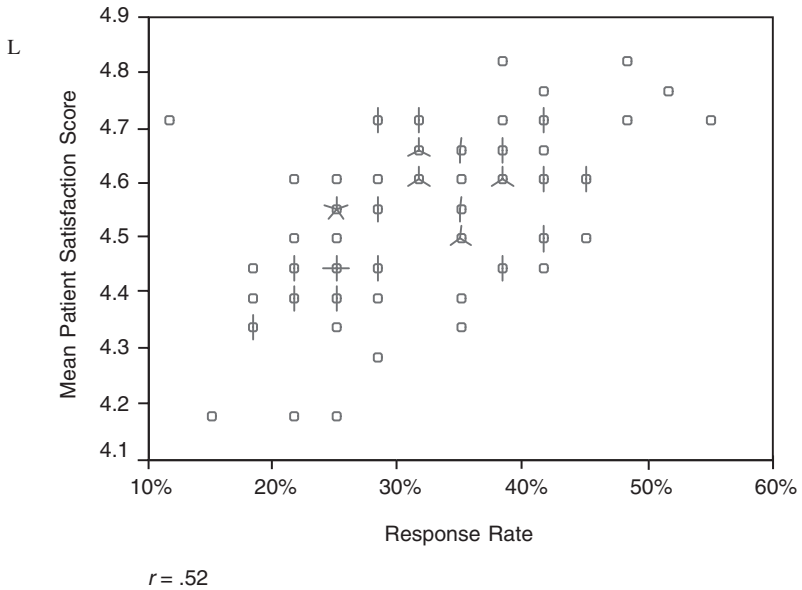
The items on the questionnaire were standard patient satisfaction questions from an operational survey implemented as part of ongoing organizational quality assurance efforts. Items addressed areas that the health care organization identified as important. Eleven questions related to satisfaction with the provider were selected for this analysis (see Appendix). Ratings were on a 5-point scale, with 1 the lowest possible rating, and 5 the highest possible rating.

For the second phase of this study, data were generated to simulate one hundred patient satisfaction ratings (on a 1 to 5 scale) for each of one hundred physicians. The goal of the simulation was to produce realistic data, comparable to the observed data, but with a known underlying distribution.

We constructed a stochastic model using the assumption of a linear relationship between the characteristics we were modeling. Preliminary analysis of the real dataset provided evidence of the appropriateness of a linear model in this case (see Figure 1). The first step in the simulation was to generate satisfaction scores for one hundred simulated physicians. These scores were normally distributed. With these scores as a starting point, one hundred patient ratings were generated for each physician, allowing differences between patients of the same provider. This procedure simulated differences between patients by allowing "easier" and "harder" raters, but did not explicitly constrain simulated patients to vary in a specified way. The simulated ratings were put on a 1 to 5 metric, comparable to the actual patient satisfaction scores. The mean was near the high end of the scale, as is typical of patient satisfaction ratings in general, and of the actual dataset analyzed here in particular.

From this simulated population of one hundred ratings per physician, two different samples were selected: a random sample, and a biased sample. In order to select a sample comparable to the observed sample, and simulate the bias posited to underlie the observed data, simulated patients were selected to be "respondents" for the biased sample in such a way that the more satisfied a patient was, the more likely that patient was to be included. The probabilities used to differentially select simulated patients as respondents were determined using an iterative process, working backwards from the real dataset until the biased sample matched the actual data in terms of response rate, satisfaction mean, and standard deviation, and the correlation between the two. It was considered desirable that the simulated data be comparable to the observed data in order that the simulation produce realistic data.

Figure 1: Mean Satisfaction Rating by Response Rate by Provider. A circle represents one provider. If there is more than one provider at a single point, the number of lines extending out from the center of the circle equals the total number of providers at that point



This data generation model and sampling procedure resulted in data that closely approximated the real patient satisfaction data. Once this model was determined, one hundred replications were conducted resulting in a total of simulated ratings for ten thousand physicians, each rated by one hundred patients.

ANALYSIS

For the actual (observed) patient satisfaction data, the response rate for each provider was calculated from the number of returned questionnaires divided by the total number of questionnaires sent. Satisfaction scores were calculated for all physicians with more than one completed patient satisfaction questionnaire, if the patient had responded to at least 10 of the 11 items identified for this study (thus the number of usable surveys was less than the number of returned surveys).

A generalizability study was conducted to estimate the generalizability of the satisfaction ratings at the provider level. In this case, a generalizability analysis was considered more appropriate than classical test theory-based estimates of reliability such as Cronbach's alpha, as the former allows assessment of multiple sources of error simultaneously, for example, both items and patients (raters). Generalizability analyses produce estimates of the variance components associated with each source of error, and also allow computation of a generalizability coefficient (g), which is a reliability coefficient comparable to alpha. Finally, the variance components from the generalizability analysis can be used to estimate reliability for different numbers of patients and items, in a manner similar to using the Spearman-Brown formula with classical reliability estimates.

The correlation between mean patient satisfaction scores and response rate was calculated. For the simulated data, three satisfaction ratings were calculated for each physician. The first, referred to as the "Full Mean," was the mean of all one hundred simulated patient ratings, without sampling. The second, referred to as the "Random Mean," was calculated based on a random sample of the patient ratings. The third, referred to as the "Biased Mean," was the mean of the sample of patient ratings that resulted from differentially sampling patients as described above, where the most satisfied patients had a greater probability of being included. For each physician, the difference between these three mean ratings was calculated.

RESULTS

The dataset of actual patient satisfaction ratings contained ratings of 82 physicians by 6,681 patients, with an average of 81 patients rating each physician (the range was 14 to 158). Descriptive statistics are reported in Table 1. The overall response rate was 32 percent; response rates for individual providers ranged from 11 to 55 percent.

Variance components estimates from the generalizability study revealed the following percentages of variance associated with each facet: provider 3 percent; patient (nested within provider) 65 percent; item 1 percent; patient by item interaction .4 percent; error 31 percent. These results make clear that varying the number of items is likely to have minimal impact on the generalizability estimate—that is, changing the number of items would have almost no impact on the reliability of the scores. Therefore, estimates of the

Table 1: Descriptive Statistics for Real Observed and Simulated Data

<i>Data Source</i>	<i>Mean</i>	<i>SD</i>	<i>Correlation between Response Rate and Satisfaction Rating</i>	<i>Response Rate</i>
Observed data	4.54	.14	.52	32%
Simulated full sample	4.41	.14	NA	100%
Simulated random sample	4.41	.15	-.01	31%
Simulated biased sample	4.53	.13	.55	32%

g-coefficient were calculated holding the number of items constant at 11, but varying the number of patients providing ratings. It was found that a g-coefficient of .80 or greater would require approximately one hundred patient raters per provider. If only 10 patient raters were available per provider, the g-coefficient would be .31; with 50 patients the reliability would increase to .69; and with one hundred patients it would increase to .81. The average number of respondents per provider in this data set was 81; with 81 respondents and 11 items the g-coefficient is estimated to be .78.

The correlation between response rate and mean satisfaction rating was .52, which is statistically significant ($p < .01$). Figure 1 displays response rate by mean satisfaction rating for each provider.

Means and standard deviations of the simulated satisfaction data are also presented in Table 1. The simulated data generation and biased sampling procedure produced data that closely resembled the real dataset (see Table 1), as was our intent. Considering the simulated data only, the mean for the random sample was identical to the mean for the full sample, but the difference between the biased sample mean and the full mean was .12, or just less than a full standard deviation.

Table 2 contains the primary results of the simulation. As expected, differences were found between the mean ratings based on the entire sample and the mean ratings after sampling. Differences were not uniform across all satisfaction levels; discrepancies were greater for those physicians with lower true satisfaction ratings. Figure 2 provides a graphic summary of typical results and highlights that differences between the biased mean and the full mean varied across physicians. Differences were greatest when true satisfaction scores were lowest.

Figure 2: Differences between Full Sample Mean Ratings and Biased Mean Ratings by Provider

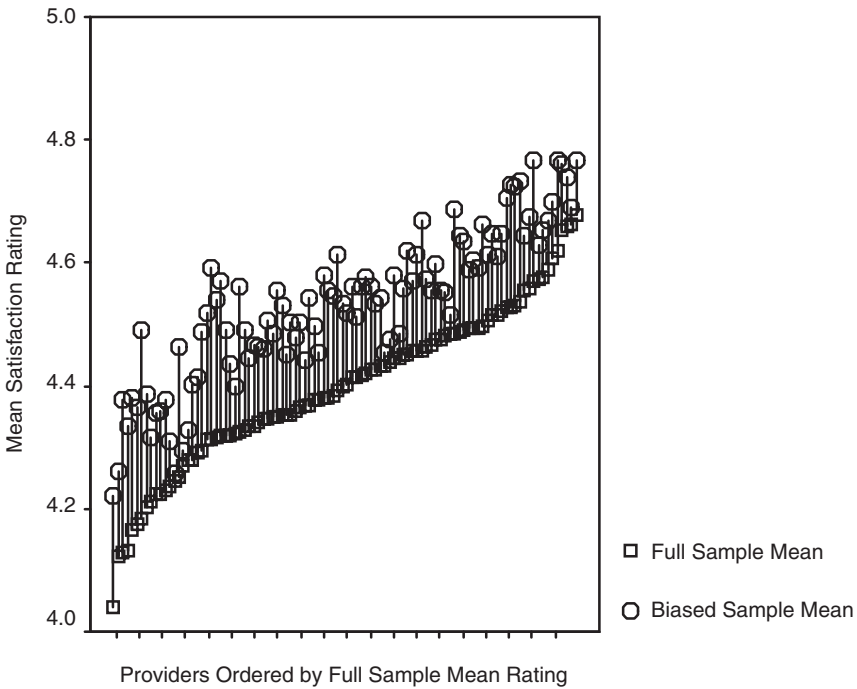


Table 2: Average Difference between Biased Sample Mean and Full Mean

<i>Physician Standing Based on Full Mean</i>	<i>Difference</i>
Bottom quartile	.16
Second quartile	.13
Third quartile	.11
Top quartile	.09
Overall	.12

Note: In all cases the mean ratings after sampling were higher than the mean ratings based on the full sample.

DISCUSSION

The results of the generalizability study demonstrate that for the observed satisfaction data, a high percentage of the variation in scores between

providers is associated with differences in patients' ratings, and a very small percentage is associated with different items. This means that while a given patient is likely to provide similar ratings across a number of items referring to a given provider, different patients rating the same provider are likely to give different ratings. This finding highlights that the score for any physician will depend to a large extent on how many patients, and which patients, provide ratings.

The relatively high correlation between response rate and mean patient satisfaction rating in the real dataset analyzed here suggests that in this instance more satisfied patients were more likely to respond than those who were less satisfied. This finding is consistent with the findings of other studies of patient satisfaction (Barkley and Furse 1996; Etter, Perneger, and Rougemont 1996; Lasek et al. 1997; Pearson and Maier 1995; Woolliscroft et al. 1994).

The results of our simulation study demonstrated that if response bias is present, it will have a meaningful impact on the results of patient satisfaction surveys. If less-satisfied patients are less likely to respond, patient satisfaction will be overestimated overall and the magnitude of the error will be greatest for physicians with the lowest patient satisfaction. For physicians who are "better" at satisfying patients, a high percentage of patients may be likely to respond and provide high ratings; for physicians with less-satisfied patients, a smaller percentage of patients will be likely to respond, and further, these respondents may be the *most* satisfied of the low-satisfaction physician's patients. This results in a bias in satisfaction scores for low-satisfaction physicians, inflating their scores relative to those of high-satisfaction physicians, thereby minimizing differences between the two. Thus, for both high- and low-satisfaction physicians, the most satisfied patients will be most likely to respond, but the difference between respondents and nonrespondents (and therefore the difference between true scores and observed scores) is likely to be greater for low satisfaction physicians than for high satisfaction physicians.

While the magnitude of the difference between the full-sample satisfaction ratings and the biased-sample satisfaction ratings may seem small (.12 across all simulations), in fact this difference is close to a full standard deviation. In addition, the difference in mean ratings for the full sample compared to the biased sample was almost twice as large for those physicians in the lowest satisfaction quartile (.16), as compared to those in the highest quartile (.09).

It is important to be clear on the impact of bias compared to random error. Both random error and bias may result in changes in relative rankings. Random error may add to or subtract from a provider's true score, and if the effect is in one direction (increase) for one provider, and in the opposite direction (decrease) for another provider, and if the true scores for these two providers are relatively close, then the rank of their observed scores may reverse as a result. The extent of the effect, and the resultant impact on relative rank, depend on the magnitude of the error relative to the variance in true scores of the providers. Even if there is no bias in the scores, if only a small number of patients respond, the magnitude of the error may be relatively large.

While random error introduces "noise" in the manner described above, bias could mask positive changes in scores, as the magnitude of the bias is a function of satisfaction. An example may help illustrate this. Imagine that provider A has relatively low true satisfaction ratings at time one. Now imagine that provider A changes, so that at time two his patients are in fact more satisfied. More patients will be likely to respond to the survey, and his "observed" score will be higher. However, at the same time, his observed score will be likely to be more accurate, and therefore the difference between his observed score and his true score will be less. Thus, provider A's observed score will increase *less* than his true score.

Our results provide an illustration of the impact of response bias under one set of realistic conditions. However, it is important to note that while we believe our simulation was realistic, other circumstances are likely to be encountered in practice. For instance, the actual survey studied here had a relatively low response rate overall, which was likely due at least in part to the absence of follow-up procedures. Surveys that do include follow-up procedures are likely to yield higher response rates, and increases in response rates are likely to reduce the impact of bias. In addition, the magnitude of the bias may also vary depending on circumstances, and we do not know at present whether the magnitude of the bias investigated here could be considered typical.

There is another important issue to consider with respect to the number of respondents. Other things being equal, fewer respondents will result in larger standard errors for satisfaction estimates for providers. In the case of patient satisfaction data analyzed at the level of the individual provider, it is possible to construct confidence intervals for each provider based on the number of patients providing ratings for that provider, so that the width of the intervals for low response-rate providers would be greater than that for high response rate providers. Used appropriately, such confidence intervals would discourage unwarranted conclusions about differences between providers, or

between an individual provider and a set standard. However, it is important to note that response bias serves to change the mean of a distribution of a set of scores, rather than simply reduce the precision of measurement. Thus, while it is advisable to take standard errors into account when making comparisons, they will not correct for biased scores.

It is important to note the limitations of this study. First, evidence of a positive relationship between response rate and satisfaction ratings in the real data analysis was based on data on primary care providers working in the context of a single health care organization. Clearly, additional research is needed to determine whether this relationship is typical of data from other organizations, other parts of the country, and other types of providers. A second limitation is the fact that the patient satisfaction questionnaire used in this study was anonymous, and therefore we were unable to link patient responses with patient characteristics.

It is also important to highlight that the simulation results are limited to the extent that the models used are reflective of what is likely to occur with real data. The model used to generate the simulated patient satisfaction ratings produced variability between providers and between patients, but did not explicitly model other factors (beyond provider and patient facets) that might produce this variability. The simulated dataset therefore is consistent with the assumption that some providers are “better” at satisfying patients than others, and that different patients are likely to give differing ratings of the same provider. We did not attempt to model the complex relationships between satisfaction and the numerous factors that may influence satisfaction, such as differences in experience with the provider, the medical issue involved, patient expectations, differences in interpretations of the items and the scale, and differences in provider characteristics including race/ethnicity, language, gender, and age. With respect to selecting respondents to simulate a biased sample, level of satisfaction was the only variable considered in determining likelihood of responding. There are almost certainly nonrandom factors that contribute to likelihood of response, and to the extent that these mitigate the relationship between satisfaction and response rate, our model is an oversimplification. While the simplicity of our model may be a limitation, our results highlight a simple but important point—if likelihood of responding is related to satisfaction, then results will be biased, regardless of what factors influence satisfaction. For example, if certain providers tend to receive lower ratings, and they have sicker patients, this does not invalidate our argument or our findings. In fact, the impact of health status may be *underestimated* if the least satisfied of the sick may be least likely to respond.

CONCLUSION

In light of the results of this study, and the limitations discussed above, it is clear that further research in this area is needed. We have provided a demonstration of how response bias attributable to differences in satisfaction could impact validity. However, in actual datasets, numerous factors may influence both satisfaction and response likelihood. Further research is needed to generate and test more complex models. Our research highlights the importance of considering not only the factors that influence satisfaction, but also response biases that serve as filters and thereby influence obtained satisfaction ratings. Additional research is needed to determine whether evidence of response bias due to differences in satisfaction exists in other datasets, and to examine possible interactions between satisfaction, response likelihood, and patient and provider characteristics.

In spite of the exploratory nature of this study, our findings do have practical implications. First, health care organization administrators and others who use patient satisfaction surveys should be aware that response biases may impact the results of such surveys, giving the impression that patients are more satisfied than they in fact are. If results are used to evaluate and compare individual providers, providers who are better at satisfying patients may be disadvantaged relative to less-satisfying peers. Not only may the relative rankings of providers change, as may happen even with random sampling error, but estimates of the magnitude of differences between providers will also be influenced. This bias could systematically deflate estimates of positive change and mask signs of improvement. Researchers working in applied settings can look for evidence of response bias by instituting follow-up procedures, and comparing early responders to more reluctant responders, although even lack of differences between these two groups will not be definitive evidence of lack of bias. Organizations should continue to strive to maximize response rates, as the impact of such biases will be minimized at higher response rates.

In summary, our findings raise concerns about the possibility of response bias in patient satisfaction surveys. Our analysis of actual patient satisfaction data suggests that the most satisfied patients may be the most likely to respond, and our simulation demonstrated how such a response bias might jeopardize the validity of interpretations based on a biased sample. Further research is needed to determine whether the effect demonstrated here is more than a theoretical possibility.

APPENDIX

Questionnaire Items with Item Means and Standard Deviations

Item	Mean (SD)	Minimum Maximum
My provider is concerned for me as a person.	4.57 (.16)	4.17 4.92
My provider spends enough time with me.	4.51 (.15)	4.23 4.89
My provider listens carefully to what I am saying.	4.58 (.14)	4.28 4.89
My provider explains my condition to my satisfaction.	4.53 (.14)	4.19 4.82
My provider explains medication to my satisfaction.	4.50 (.15)	4.08 4.81
My provider supplies me with the results of my tests in a timely fashion.	4.46 (.19)	3.76 4.78
My provider supplies information so I can make decisions regarding my own care.	4.46 (.16)	4.06 4.79
My provider refers me to specialists as needed.	4.53 (.15)	4.00 4.77
My provider treats me with respect and courtesy.	4.70 (.11)	4.44 4.89
My provider returns my telephone calls within a reasonable period of time.	4.49 (.18)	3.75 4.77
I would recommend my provider to family and friends.	4.61 (.15)	4.23 4.86

Number of providers = 82.

All responses were on a 5-point scale with 1 = Strongly Disagree and 5 = Strongly Agree.

REFERENCES

- Asch, D., M. Jedziewski, and N. Christakis. 1997. "Response Rates to Mail Surveys Published in Medical Journals." *Clinical Epidemiology* 50 (10): 1129-36.
- Band, P., J. Spinelli, W. Threlfall, R. Fang, N. Le, and R. Gallagher. 1999. "Identification of Occupational Cancer Risks in British Columbia. Part 1: Methodology, Descriptive Results, and Analysis of Cancer Risks, by Cigarette Smoking Categories of 15,463 Incident Cancer Cases." *Journal of Occupational and Environmental Medicine* 41 (4): 224-32.
- Barkley, W. M., and D. H. Furse. 1996. "Changing Priorities for Improvement: The Impact of Low Response Rates in Patient Satisfaction." *Joint Commission Journal on Quality Improvement* 22 (6): 427-33.

- Benfante, R., D. Reed, C. MacLean, and A. Kagan. 1989. "Response Bias in the Honolulu Heart Program." *American Journal of Epidemiology* 130 (6): 1088–100.
- Carlson, M. J., J. Blustein, N. Fiorentino, and F. Prestianni. 2000. "Socioeconomic Status and Dissatisfaction Among HMO Enrollees." *Medical Care* 38 (5): 508–16.
- Diehr, P., T. Koepsell, A. Cheadle, and B. Psaty. 1992. "Assessing Response Bias in Random-Digit Dialing Surveys: The Telephone-Prefix Method." *Statistics in Medicine* 11 (8): 1009–21.
- Etter, J., and T. Perneger. 1997. "Analysis of Non-Response Bias in a Mailed Health Survey." *Journal of Clinical Epidemiology* 50 (10): 1123–8.
- Etter, J. F., T. V. Perneger, and A. Rougemont. 1996. "Does Sponsorship Matter in Patient Satisfaction Surveys? A Randomized Trial." *Medical Care* 34 (4): 327–35.
- Ford, R. C., S. A. Bach, and M. D. Fottler. 1997. "Methods of Measuring Patient Satisfaction in Health Care Organizations." *Health Care Management Review* 22 (2): 74–89.
- Gold, M., R. Hurley, T. Lake, T. Ensor, and R. Berenson. 1995. "A National Survey of the Arrangements Managed-Care Plans Make with Physicians." *New England Journal of Medicine* 333 (25): 1678–83.
- Goodfellow, M., N. E. Kiernan, F. Ahern, and M. A. Smyer. 1988. "Response Bias Using Two-Stage Data Collection: A Study of Elderly Participants in a Program." *Evaluation Review* 12 (6): 638–54.
- Heilbrun, L. K., P. D. Ross, R. D. Wasnich, K. Yano, and J. M. Vogel. 1991. "Characteristics of Respondents and Nonrespondents in a Prospective Study of Osteoporosis." *Journal of Clinical Epidemiology* 44 (3): 233–9.
- Hill, A., J. Roberts, P. Ewings, and D. Gunnell. 1997. "Non-Response Bias in a Lifestyle Survey." *Journal of Public Health Medicine* 19 (2): 203–7.
- Hoeymans, N., E. Feskens, G. V. D. Bos, and D. Kromhout. 1998. "Non-Response Bias in a Study of Cardiovascular Diseases, Functional Status and Self-Rated Health Among Elderly Men." *Age and Ageing* 27 (1): 35–40.
- Jay, G. M., J. Liang, X. Liu, and H. Sugisawa. 1993. "Patterns of Nonresponse in a National Survey of Elderly Japanese." *Journal of Gerontology* 48 (3): S143–52.
- Krosnick, J. A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537–67.
- Lasek, R. J., W. Barkley, D. L. Harper, and G. E. Rosenthal. 1997. "An Evaluation of the Impact of Nonresponse Bias on Patient Satisfaction Surveys." *Medical Care* 35 (6): 646–52.
- Launer, L. J., A. W. Wind, and D. J. H. Deeg. 1994. "Nonresponse Pattern and Bias in a Community-Based Cross-Sectional Study of Cognitive Functioning among the Elderly." *American Journal of Epidemiology* 139 (8): 803–12.
- Livingston, P., S. Lee, C. McCarty, and H. Taylor. 1997. "A Comparison of Participants with Non-Participants in a Population-Based Epidemiologic Study: The Melbourne Visual Impairment Project." *Ophthalmic Epidemiology* 4 (2): 73–81.
- Macera, C., K. Jackson, D. Davis, J. Kronenfeld, and S. Blair. 1990. "Patterns of Non-Response to a Mail Survey." *Journal of Clinical Epidemiology* 43 (12): 1427–30.

- Norton, M. C., J. C. S. Breitner, K. A. Welsh, and B. W. Wyse. 1994. "Characteristics of Nonresponders in a Community Survey of the Elderly." *Journal of the American Geriatrics Society* 42: 1252-6.
- O'Neill, T., D. Marsden, A. Silman, and The European Vertebral Osteoporosis Study Group. 1995. "Differences in the Characteristics of Responders and Non-Responders in a Prevalence Survey of Vertebral Osteoporosis." *Osteoporosis International* 5 (5): 327-34.
- Panser, L., C. Chute, H. Guess, J. Larsonkeller, C. Girman, J. Oesterling, M. Lieber, and S. Jacobsen. 1994. "The Natural History of Prostatism: The Effects of Non-Response Bias." *International Journal of Epidemiology* 23 (6): 1198-205.
- Pearson, D., and M. L. Maier. 1995. "Assessing Satisfaction and Non-Response Bias in an HMO-Sponsored Employee Assistance Program." *Employee Assistance Quarterly* 10 (3): 21-34.
- Prendergast, M., J. Beal, and S. Williams. 1993. "An Investigation of Non-Response Bias by Comparison of Dental Health in 5-Year-Old Children According to Parental Response to a Questionnaire." *Community Dental Health* 10 (3): 225-34.
- Rockwood, K., P. Stolee, D. Robertson, and E. R. Shillington. 1989. "Response Bias in a Health Status Survey of Elderly People." *Age and Ageing* 18: 177-82.
- Rosenthal, G. E., and S. E. Shannon. 1997. "The Use of Patient Perceptions in the Evaluation of Health-Care Delivery Systems." *Medical Care* 35 (11, supplement): NS58-68.
- Smith, C., and D. Nutbeam. 1990. "Assessing Non-Response Bias: A Case Study from the 1985 Welsh Heart Health Survey." *Health Education Research* 5 (3): 381-6.
- Templeton, L., A. Deehan, C. Taylor, C. Drummond, and J. Strang. 1997. "Surveying General Practitioners: Does a Low Response Rate Matter?" *British Journal of General Practice* 47 (415): 91-4.
- Tennant, A., and E. Badley. 1991. "Investigating Non-Response Bias in a Survey of Disablement in the Community: Implications for Survey Methodology." *Journal of Epidemiology and Community Health* 45 (3): 247-50.
- van den Akker, M., F. Buntinx, J. Metsemakers, and J. Knottnerus. 1998. "Morbidity in Responders and Non-Responders in a Register-Based Population Survey." *Family Practice* 15 (3): 261-3.
- Vestbo, J., and F. Rasmussen. 1992. "Baseline Characteristics Are Not Sufficient Indicators of Non-Response Bias Follow-Up Studies." *Journal of Epidemiology and Community Health* 46 (6): 617-9.
- Williams, P., and A. Macdonald. 1986. "The Effect of Non-Response Bias on the Results of Two-Stage Screening Surveys of Psychiatric Disorder." *Social Psychiatry* 21 (4): 182-6.
- Woolliscroft, J. O., J. D. Howell, B. P. Patel, and D. E. Swanson. 1994. "Resident-Patient Interactions: The Humanistic Qualities of Internal Medicine Residents Assessed by Patients, Attending Physicians, Program Supervisors, and Nurses." *Academic Medicine* 69 (3): 216-24.
- Young, G. J., M. Meterko, and K. R. Desai. 2000. "Patient Satisfaction with Hospital Care: Effects of Demographic and Institutional Characteristics." *Medical Care* 38 (3): 325-34.