

Floyd Jackson Fowler, Jr., Center for Survey Research, UMass - Boston  
100 Morrissey Boulevard, Boston, MA 02125-3393

Key Words: mail survey, nonresponse, bias, CAHPS

## INTRODUCTION

Mail surveys are being widely used to collect data about the experience of health plan members. It is not uncommon for the resulting response rates to be 40 percent or lower. A key concern is whether or not the nonresponse is biased and if results would be different if response rates were higher.

The CAHPS (Consumer Assessment of Health Plans) project is funded by the Agency for Health Care Policy Research. A consortium of researchers working with RAND, the Research Triangle Institute, and the Harvard Medical School are working together to design instruments and protocols to produce comparable data about health plan experiences. The National Committee on Quality Assurance (NCQA) has a similar goal. It has developed a Member Satisfaction Survey (MSS) and sets of procedures for plans to use a survey their members. Both NCQA and CAHPS had an interest in the potential bias of nonresponse to mail surveys. The research reported here was a cooperative effort to study the issue.

## METHODS

Four managed health plans in Colorado volunteered to help with the project. Probability samples of 800 adults who had been enrolled in the plans for a least 12 months were drawn from each of the plans' member lists. A 10-page survey instrument that included the CAHPS core questions, plus a sampling of key questions from the NCQA instrument, was developed. An advance letter was sent from the plans to sampled members. A mail survey protocol, consisting of an initial mailing, a postcard, and a second mailing was carried out.

After the mail returns had stopped arriving, an effort was made to carry out telephone interviews with mail nonrespondents. The CAHPS questions are designed so wording can be virtually identical whether it is interviewer or self administered.

As Table 1 shows, the response rates to the mail solicitation alone ranged from 42 to 50 percent. After the telephone effort, the total response rates were raised to from 60 to 72 percent. For each plan, data were added from approximately 20 percent of all plan members who had not responded by mail.

An obvious issue is that the initial returns were obtained by mail, while the nonrespondents provided data to a telephone interviewer. Technically, it is not possible

to completely dissociate the effects of the data collection mode from the characteristics of the people who were interviewed by telephone. However, because of a well-controlled experiment in Washington state, in which data were collected by mail and by telephone from quite comparable samples, we have good data on which CAHPS items are and are not likely to be affected by mode of data collection. We can take advantage of that information to help sort out the likely differences that are due to nonresponse alone. The analysis that follows compares those responding by mail with those whose data were collected by telephone.

## RESULTS

### Comparing Mail Respondents with Telephone Interviews with Mail Nonrespondents

Table 2 provides a summary of results when all the items in the survey instrument were compared. It can be seen that of the 52 items from the CAHPS instrument in the test, 30 had a statistically significant difference ( $P < .05$ ) between the mail respondents and the mail nonrespondents. For the 11 items from the NCQA Member Satisfaction Survey that were included in the instrument, 5 showed a statistically significant difference. Although we do not have a good mode test of the MSS items, in the Washington test, only 9 of the CAHPS items showed a statistically significant difference between data collected by telephone and those collected by mail.

There were four major demographic differences between mail respondents and nonrespondents. Over 30 percent of the telephone respondents were under 35 years of age, compared to 20 percent of the mail respondents. Only 37 percent of the mail respondents were male, while 51 percent of the telephone respondents were male. The mail respondents were also distinctively likely to be white; Hispanics and other minorities were more likely to be found among the telephone respondents. (Table 3)

Those not responding to the mail survey were also likely to be healthier and lower utilizers of medical care. The mail respondents were 50 percent more likely to have been hospitalized in the past year than the nonrespondents; they were 50 percent more likely to have reported a chronic medical condition. Health ratings of nonrespondents were better than those of respondents. Finally, 42 percent of the nonrespondents interviewed by telephone had not been to a doctor's office in the past 6

months, while that was true for only 26 percent of the mail respondents. (Table 4)

### The Effects on Plan Ratings

Perhaps the most important question is the extent to which the overall ratings of plans are affected by nonresponse. In an effort to address that question, we looked at three key overall measures: the MSS overall satisfaction question and the CAHPS 0-10 ratings of overall health care and of the plan.

For each of these measures of quality, data are presented from the mail survey alone, as if no followup of nonrespondents by telephone had been attempted, and then for the total sample, when the telephone results are added to the initial mail returns. Differences in the two numbers indicate the extent to which accepting the response rate obtained by mail alone, without telephone interviews with nonrespondents, affected final descriptions of the plans.

In Table 5, the percentage of respondents saying they were "completely" or "very" satisfied changed by one percentage point for two plans; another plan changed by two percentage points; a third changed by three percentage points. With samples of this size, when comparing plans, a standard error of differences is about three percentage points. Hence, a change of three percentage points could clearly make a difference in the results of comparing plans.

Table 6 presents the mean ratings of health care and of the plans. For these ratings, a difference of 0.2 between plans is likely to be statistically significant. All of the ratings of the plans went up when the telephone responses are added. The comparisons of plans did not change much for the rating of health care. However, for the overall plan ratings, Plan B's rating improved to a statistically significant degree, and its standing compared to other plans improved significantly. Plan D's rating also improved by about one standard error.

### CONCLUSION

The clearest aspect of these data is that the people who can be interviewed by telephone, who did not return mail questionnaires, are quite different in some important respects. They clearly include more young adults, more males, lower users of health services, and people who think they are in better health. They also include more ethnic minorities. It may be understandable why some of these groups, such as young healthy males, think the quality of their medical care is less central to their lives than those who respond readily to mail surveys about their health care. Nonetheless, they are part of the population served, and systematically underrepresenting their views is obviously undesirable.

We have done many comparisons of the extent to which the answers and experiences of those nonrespondents who can be interviewed by telephone differ from the people who return mail questionnaires. In 35 out of 63 items, there is a statistically significant difference between the mail respondents and those mail nonrespondents who were interviewed by telephone. Only a few of these differences could reasonably be attributed to the change in the mode of data collection.

Of course, since mail respondents outnumbered the telephone interviews two to one, the answers of those responding by telephone had to be quite different in order to make a big difference in the aggregate estimates. Many estimates were not much different with and without the inclusion of the telephone responses. However, Tables 5 and 6 show that even with only four plans to compare, overall ratings of health plans did change when the telephone respondents were added in ways that significantly affect comparisons.

The critical issue is what the implications are for protocols for collecting data about people's experiences with their health plans. On the one hand, by general survey research standards, response rates below 60 or even 70 percent are often considered problematic and likely to be biased. The data from the telephone interviews show that those concerns are not unfounded. There are important differences between the telephone respondents and the mail respondents. On the grounds of credibility alone, having a more representative inclusion of people in better health, of young adults, of males, and of ethnic minorities is valuable. Moreover, it seems most likely that when there are differences, the data from combining mail returns with telephone interviews of nonrespondents are more credible and more likely to accurately reflect patient experiences overall.

**TABLE 1:  
Response Rates by Health Plan**

	<b>PLAN A</b>	<b>PLAN B</b>	<b>PLAN C</b>	<b>PLAN D</b>
Initial sample size	800	800	800	800
Ineligible respondents*	60	68	84	85
Eligible respondents	740	732	716	715
Mail returns	370	350	309	302
Percent of mail returns	50%	48%	43%	42%
Phone interviews	156	176	135	127
Percent of phone returns	21%	24%	19%	18%
<b>Total percent of respondents</b>	<b>71%</b>	<b>72%</b>	<b>62%</b>	<b>60%</b>

\* Confirmed deceased or no longer plan member. All others selected, including those for whom there was not a good address, were presumed "eligible."

**TABLE 2:  
Number of Significant Differences Between Initial Mail Responses and Telephone Interviews with Mail Nonrespondents by Source of Question and Question Type**

	SIGNIFICANTLY DIFFERENT (p<.05)		
	YES	NO	TOTAL
<b>QUESTION ORIGIN — CAHPS</b>			
YES/NO Questions	11	4	15
Always/Never Questions	8	13	21
Health Status	4	1	5
Utilization	1	1	2
Demographics	4	1	5
0 to 10 rating scales	2	2	4
<b>TOTAL</b>	<b>30</b>	<b>22</b>	<b>52</b>
<b>QUESTION ORIGIN — MSS</b>			
Excellent to Poor ratings	3	3	6
Problem questions	--	3	3
Satisfaction question	1	--	1
Wait for appointment	1	--	1
<b>TOTAL</b>	<b>5</b>	<b>6</b>	<b>11</b>

**TABLE 3:**  
**Comparison of Results from Demographic Questions that Differed Significantly (p <.05):**  
**Initial Mail Responses v. Telephone Interviews with Mail Non-Respondents**

		<b>MAIL</b>	<b>TELEPHONE</b>
<i>Question c54 (Age)</i>	18-24	4%	9%
	25-34	16	23
	35-44	27	33
	45-54	29	26
	55-64	20	8
	65-74	4	1
<i>Question c55 (Gender)</i>	Male	37%	51%
	Female	63	49
<i>Question c57 (Hispanic)</i>	Yes	11%	17%
	No	89	83
<i>Question c58 (Race)</i>	Native American	1%	1%
	Asian	1	1
	Black	3	5
	White	87	76
	Other	8	17

**TABLE 4: Comparison of Results from Health Status Questions that Differed Significantly (p <.05): Initial Mail Responses v. Telephone Interviews with Mail Non-Respondents**

		MAIL	TELEPHONE
<i>Question c49 (Hospital overnight - 12 mos)</i>	Yes	12%	8%
	No	88	92
<i>Question c50 (Medical condition 3+ mos)</i>	Yes	46%	30%
	No	54	70
<i>Question c51 (Seen dr 2+ times - 12 mos)</i>	Yes	73%	61%
	No	27	39
<i>Question c53 (Rate overall health)</i>	Excellent	19%	27%
	Very good	40	42
	Good	32	24
	Fair	8	6
	Poor	1	1

**TABLE 5: MSS Satisfaction Measure by Health Plan: Mail Sample v. Combined Sample (mail plus telephone)**

SATISFACTION				
<i>Question 48 (How satisfied with plan)</i>	PLAN A	PLAN B	PLAN C	PLAN D
Mail sample (% completely and very satisfied)	61	55	56	52
Combined sample (% completely and very satisfied)	62	56	58	55

**TABLE 6: Mean CAHPS Ratings by Health Plan: Mail Sample v. Combined Sample (mail plus telephone)**

Ratings (0-10 where 10 is best)				
<i>Question c37 (Health care)</i>	PLAN A	PLAN B	PLAN C	PLAN D
Mail sample	7.56 (SE = .12)	7.74 (SE = .13)	7.98 (SE = .12)	7.74 (SE = .13)
Combined sample	7.65 (SE = .10)	7.86 (SE = .11)	8.04 (SE = .10)	7.82 (SE = .11)
<i>Question 45 (Health insurance plan)</i>				
Mail sample	7.61 (SE = .11)	7.25 (SE = .12)	7.39 (SE = .12)	7.09 (SE = .13)
Combined sample	7.65 (SE = .09)	7.44 (SE = .09)	7.41 (SE = .10)	7.19 (SE = .11)