

The Nature of Nonresponse in a Medicaid Survey: Causes and Consequences

Patricia M. Gallagher¹, Floyd Jackson Fowler Jr.¹, and Vickie L. Stringfellow¹

This study examines the effect of multiple phases of data collection on 1) the representativeness of the respondents and 2) survey responses to substantive items. A version of the CAHPS survey instrument was used to collect information from 1,600 adults covered by Massachusetts Medicaid about their experiences with getting health care through their health plans. In order to gain a better understanding of the significance of nonresponse, intensive follow-up with nonrespondents was done utilizing three phases of data collection: mail, telephone, and in-person protocols in successive waves of contact. Using administrative data, we were able to compare demographic and health characteristics of respondents to each phase, as well as cumulatively, with the characteristics of the total sample.

The results indicate that the addition of completed surveys from each successive phase of data collection yielded a final group of respondents that is more representative of the total population with respect to age, race, presence of a chronic condition, and annual medical costs. Responses to 40 key survey items were also analyzed by phase and cumulatively. Responses from mail and telephone respondents were significantly different for eleven of these items; responses collected from in-person interviews were significantly different from the combined data from mail and telephone respondents for three of these items.

Key words: Nonresponse; bias; Medicaid; mail; telephone; in-person interview; face to face interview; CAHPS.

1. Introduction

Medicaid programs throughout the United States are surveying members to evaluate the quality of their medical care experiences. These surveys are most often done by mail, sometimes with telephone follow-up of mail nonrespondents to increase response rates. Even in mixed mode surveys of this population, response rates are often less than satisfactory. The response rates for mail surveys of this type typically range from 20% to 50% (Farley and Jael 2000). While telephone efforts usually add another 5 to 15 percentage points to these rates, these efforts increase the cost and complexity of survey administration. One obvious question is how much the telephone effort to reduce mail nonresponse affects the results. Another is the extent to which the data from the combination of mail and telephone protocols are representative of the entire population. Adding in-person interview

¹ University of Massachusetts Boston, Center for Survey Research, 100 Morrissey Boulevard, Boston, MA 02125, U.S.A. Email: patricia.gallagher@umb.edu

Acknowledgments: The research reported here was supported by a cooperative agreement (Award #AHRQ U18HS09205) from the Agency for Healthcare Research and Quality with the Department of Health Care Policy, Harvard Medical School. We also gratefully acknowledge the assistance and cooperation of the Division of Medical Assistance, Commonwealth of Massachusetts, and the MassHealth members who kindly responded to our survey. An earlier version of this article was presented at the International Conference on Survey Nonresponse, Portland, Oregon, October 28–31, 1999.

of nonrespondents entails even higher per case costs (Hox and deLeeuw 1994; Teitler, Reichman, and Sprachman 2003).

When respondents are systematically different from nonrespondents in characteristics that are germane to the research questions being addressed by the survey, nonresponse bias occurs (Groves and Lyberg 1988; Groves 1989). Previous studies have found that respondents and nonrespondents to mailed surveys often differ in their demographic characteristics (Fowler et al. 2002). A problem for methodologists is that information on nonrespondents is often limited. When data are available, studies have found that nonrespondents to health-related mail surveys are more likely to have less than a high school education and to be male, younger, nonwhite, and unmarried than respondents (McHorney et al. 1994; Lasek et al. 1997).

The saliency of the survey topic can also differentially influence response by mode. In mail surveys, potential respondents have the opportunity to review the entire instrument before deciding whether to respond. Returns to mail surveys may be related to survey content and hence potentially biased. In contrast, nonresponse to telephone surveys may be less directly related to survey content because respondents only have a general idea about what will be asked. However, Groves, Singer, and Corning (2000) found that those with high levels of community involvement responded to a telephone survey without incentives being offered more than those less involved. From a mixed mode survey of health plan members, there is evidence that the generalizability of mail returns is limited with respect to demographic characteristics, health status, and utilization of health services. Respondents to the mail mode were less likely to be younger (under age 35), male, Hispanic, and white than respondents to the telephone mode. In addition, mail respondents were more likely to report that they were in poorer health, had a chronic condition, had been hospitalized in the last year, and had utilized more medical care than those interviewed on the phone (Fowler et al. 2002).

In order to gain a better understanding of the significance of the nonresponse, we conducted a survey of Medicaid recipients that involved intensive follow-up with three phases of data collection: mail, telephone, and in-person protocols in successive waves of contact. The survey used a CAHPS (Consumer Assessment of Health Plans Study) instrument to measure enrollees' health care experiences. Using information from the sampling frame and other administrative data, we examined the characteristics of enrollees who did and did not respond to each subsequent wave of data collection.

A critical feature of the study is that Medicaid claims data were available for the entire sample for the two years prior to data collection. Using data from both administrative records and survey responses, we were able to address three research questions. First, do nonrespondents to successive phases of contact differ systematically from respondents to previous phases? Second, does increasing response rates using multiple data collection protocols yield data that are more representative of the sample population? Third, how does improving response rates affect the estimates based on the survey?

2. Methods

2.1. Data collection protocols

The study design called for three successive phases of data collection: mail, telephone, and in-person.

2.1.1. Mail phase

The first data collection step was to mail all selected individuals a questionnaire and a fact sheet with answers to frequently asked questions. Seven to ten days later a reminder/thank you postcard was sent to the entire sample. About two weeks after the initial mailing, a replacement questionnaire packet was sent to all those who had not yet responded. Note that refusals were considered terminal noninterviews and were not contacted again.

2.1.2. Telephone phase

Approximately four weeks after the initial mailing, efforts were made to interview all nonrespondents for whom we had a phone number (obtained from Medicaid records, a computerized telephone number look-up service, or calls to Directory Assistance). A minimum of six calls were placed at different times of the day and on different days of the week, with daytime and evening call attempts occurring midweek and on weekends. In addition to new refusals, members who were away for the duration of the study, members who were too ill to participate and members who did not speak English or Spanish were considered terminal noninterviews and were not contacted again.

2.1.3. In-person phase

Nonrespondents to the mail data collection phase for whom no telephone number could be identified were immediately assigned to the in-person phase. Additionally, all nonrespondents to the telephone data collection efforts were transferred to the in-person phase where field interviewers tried to contact nonrespondents at their last known addresses. Interviewers were required to make a minimum of six attempts, including Saturday and evening visits, over at least a two-week period. The interviewers went to extraordinary lengths to track nonrespondents, examining town lists, contacting landlords, and talking to neighbors in an effort to find all sample members. Those contacted were usually interviewed face-to-face, but interviewers were allowed to offer the option of completing a self-administered form. The interviewers were also permitted to conduct telephone interviews for cases where in-person visits produced a telephone number but the designated respondent was not available when they went to the address.

2.2. *Language*

Because a substantial portion of those receiving Medicaid benefits in Massachusetts speak Spanish as their primary language, respondents had the option of completing the survey in either Spanish or English at each phase of data collection. The survey was not offered in any other language.

2.3. *Sample*

The sampling frame was provided by the Division of Medical Assistance (DMA) which oversees the administration of Medicaid in Massachusetts. In order to qualify for the study, members had to be aged 18 to 64 and continuously enrolled in the same health plan for at least 136 days in the previous six months, with no breaks in enrollment exceeding 45 days. Data for this study were collected between May and October, 1998.

The demographics of the adult Medicaid population in Massachusetts differ from those of the general United States population in some respects. More women, younger adults, and people with lower than average income receive Medicaid benefits and are thus over-represented in this sample. For example, 25% of the sample is male; 49% is under age 35. Regarding race and ethnicity, the sample is more similar to the general population overall: 68% are White (non-Hispanic), another 13% are Black (non-Hispanic), another 15% are Hispanic, and 4% are in another race/ethnicity category.

Two samples of members were drawn from a single health plan. First, a probability sample of 800 adults who qualified for federal or state level Supplemental Security Income (SSI) was drawn. A second sample of 800 adults was then drawn from the remainder of the eligible population in the same plan.

2.4. Instrument design

The CAHPS survey instruments were developed to collect information from health plan members about their experiences with getting health care through their health plans. Instrument development included multiple rounds of focus groups, cognitive interviews, and field pilot tests (Weinberger 1999). A critical feature of the CAHPS[®] instruments is that they were designed to produce comparable data whether interviewer- or self-administered. Item wording is virtually identical for mail and interviewer-administered versions of the instrument. Technically, it is not possible to completely dissociate the effects of mode of data collection from the characteristics of the people who responded at each phase of data collection. However, because of well-controlled experiments in Washington state and California in which samples of privately and publicly insured health plan enrollees were randomized to mail and telephone modes of administration, we have good data on the extent to which CAHPS items are and are not likely to be affected by mode of data collection (Fowler et al. 1999). When differences by mode were found in two rounds of testing, items were revised and retested, largely reducing mode effects. We can take advantage of that information to help sort out the likely differences that are due to nonresponse alone.

2.5. Additional administrative data

Several demographic characteristics (e.g., age, race, and gender) for each sampled member were obtained from DMA's administrative database. Additional information about diagnoses and insurance claims was also provided. These data allowed us to examine differences among those who responded at each phase of data collection and those who did not respond at all.

3. Analysis Plan

The overall goal was to describe the effects of multiple phases of data collection on response rates, the reasons for nonresponse, the characteristics of those who did and did not respond, and how these differences affect the survey responses. To do this, we carried out four sets of analyses.

3.1. *Response rates*

The goal of the first set of analyses was to examine the effect on response rates of offering mail nonrespondents first the opportunity to respond by telephone, and subsequently, in-person interviews at home. Response rates and case dispositions were calculated both by phase of contact and cumulatively.

3.2. *Differences in self-reported characteristics*

The goals of the second set of analyses were to determine whether self-reported characteristics of respondents varied at each phase of the study and cumulatively. To do this, the proportion of respondents reporting certain characteristics are presented by phase of data collection. In addition, the responses to these items were compared: first, mail versus telephone responses, and then mail plus telephone versus responses obtained during the field phase.

To adjust for the disproportionate sampling in the SSI group, in these and all subsequent analyses, weights were applied to adjust for the difference in the probability of selection.

3.3. *Administrative data for respondents and nonrespondents*

The goals of the third set of analyses were to use administrative data to assess how much additional phases of data collection affected sample representativeness. Administrative records for the total sample were compared both with the cumulative results for respondents from each phase of data collection and with the results for those who never responded.

3.4. *Effects of extensive tracking on key survey results*

The goal of the fourth set of analyses was to determine if improving response rates by offering nonrespondents multiple ways to respond affects key survey results. To accomplish this, respondents' global ratings of various aspects of their care were compared across the three phases of administration. In addition, survey items were grouped by response task and compared to see if there were differences in the way people responded by phase of data collection. These response task groups included: questions about interactions with health care providers using a four-point scale from "never" to "always;" questions about access to care using a three-point scale from "big problem" to "not a problem;" global ratings using an eleven-point scale from 0 to 10; screening questions to identify those who needed or used services using a "yes/no" response task; and questions about the number of visits to doctors' offices and emergency rooms.

4. **Results**

4.1. *Response rates by phase of contact*

In this article, we use the term "phase" rather than "mode" advisedly. The study was not designed to examine effects of mode of administration, but to explore the implications of multiple contacts with nonrespondents, employing successive modes of administration.

Here, the term “phase” signifies at which stage of contact surveys were completed rather than mode of data collection. During the telephone phase there were a few respondents who mailed their completed questionnaires, most often as a result of the telephone prompts. Similarly, a small minority of responses collected in the in-person, or “field” phase, were self-administered or collected by telephone. These responses would not have been collected via any mode if the cases had not been exposed to the phase of data collection during which they were obtained. Analyses present returns by phase, which indicate returns from all modes.

Table 1 presents cumulative response rates by phase of data collection. The response rates were calculated as the proportion of the eligible sample responding (using Equation RR1 from AAPOR (2000)), and sample members with incorrect contact information are assumed eligible. Predictably, each successive contact improved response rates. However, many cases could not be successfully contacted in the telephone phase because telephone numbers were not available. Therefore, as can be seen from the percentage of the sample that responded during each phase (see Table 1), the telephone protocol was the least productive of the three phases employed.

Overall, about 36% of the eligible sample responded during the mail phase, another 11% was picked up by adding the telephone phase, and the in-person effort added 20%, bringing the cumulative response rate to 68%.

Only about 6% of the total sample refused, and refusal conversions were not attempted. A proxy or respondent refusal at any stage was considered a final outcome. It could be argued that members for whom no final result was obtained could be considered unwilling respondents who had the chance to respond but tacitly refused by avoiding contact with interviewers. Even if one assumes that this was the case for every member for whom no

Table 1. Response and cooperation rates cumulatively and at each phase of data collection

	Mail	Phone	Field	Total
Starting sample	1,600	1,019	719	1,600
Ineligible	2	21	25	48
Completed survey	36% (572)	17%† (171)	46%‡ (317)	68% (1,060)
Refusal	*	6%	4%	6%
Inadequate contact information	12%	66%	38%	17%
Other noninterview††	–	5%	5%	5%
No final result	52%	6%	7%	3%
	100% (1,598)	100% (998)	100% (694)	100% (1,552)
% of total sample responding during phase	36%	11%	20%	–
Cumulative response rate	36%	47%	68%	68%
Cooperation rate after each phase	41%	51%	74%	82%

*Less than 0.5%

†Includes mail surveys received during this phase of contact

‡Includes mail surveys received and phone interviews completed during this phase of contact

††Includes members who were away for the duration of the study, members who were too ill to participate in the study, and members who did not speak English or Spanish

final result could be achieved, this group accounts for only a small piece (3%) of the total nonresponse (see Table 1).

Incomplete and incorrect contact information contributed to nonresponse. It can be seen in Table 1 that 12% of the mail surveys were undeliverable because of incorrect mailing addresses, and there may have been more that never reached the sampled person that were not returned. The quality of telephone contact information was the worst problem. Working phone numbers could not be obtained for 66% of those assigned to the telephone phase. Even with the intensive in-person data collection effort, 17% of the sample overall could not be located in any phase with the information available.

Another way to think about data collection results is to calculate the rate of cooperation. This is the proportion of all eligible sample members who were contacted at all and who responded. The cooperation rate allows an estimation of respondents' willingness to participate that is independent of the quality of contact information provided. While the overall cooperation rate for the three protocols was 82% (using Equation COOP1, AAPOR (2000)), there was considerable variation by phase. As the last row of Table 1 shows, 41% of those who received a mail questionnaire returned it, and about half of those reached by phone were interviewed. However, 74% of those contacted in person responded. Particularly since most of those assigned for an in-person visit had prior opportunities to respond, it is clear that the in-person contact was the most effective way to get a response.

A review of respondent tracking sheets for cases assigned to in-person interviewers indicates that only a handful of addresses were not residential; these turned out to be group homes, half-way houses, businesses, or Department of Social Services offices. Extensive efforts were made to locate sample members. However, for 38% of those assigned to the field phase, interviewers were never able to find a correct address for the enrollee. For some nonrespondents, the reason for nonresponse was a practical one; 2% of the total sample was too ill to respond, and about 3% did not speak either English or Spanish (data not presented).

4.2. *Self-reported respondent characteristics*

Table 2 compares the reported characteristics of respondents by the phase in which they responded. Comparing mail and phone respondents, it can be seen that respondents who self-report being in fair or poor health were more likely than others in better health to respond by mail, as were sampled individuals who were 35 or older. Because the mail form was in dual-language format, Spanish speakers were distinctively likely to respond by mail as well.

Different people respond at each phase and each mode has strengths. A given mode of administration may be more effective with certain subgroups. Offering a telephone interview to mail nonrespondents tended to increase responses from younger people, those whose primary language is English, and those with higher levels of education. At the same time, people who reported speaking a language other than English were far less likely to be interviewed over the telephone than they were to respond to mail and in-person contacts.

Table 2. Self-reported characteristics of adult respondents to each phase of study^a (%(n))

	Mail	Phone	Field
Rating of overall health			
Excellent	11	15	18
Very good	19	24	20
Good	29	31	31
Fair or poor	41	30*	32
Total	100% (488)	100% (152)	100% (285)
Current age			
less than 35	42	51*	57
35 or older	58	49	43
Total	100% (492)	100% (152)	100% (285)
Primary language			
English	65	82*	71**
Spanish	22	14	20
Other	13	3	9
Total	100% (492)	100% (154)	100% (285)
Office visits in last 6 months			
1 visit or more	73	77	75
No visits	27	23	25
Total	100% (427)	100% (146)	100% (279)
Education			
< High school grad	41	32	42
HS grad or GED	34	44	34
Some college or more	25	24	24
Total	100% (474)	100% (151)	100% (282)

^aData weighted to adjust for oversampling of adults in SSI or other special programs

*Mail and telephone respondents significantly different by chi-square test ($p < .05$)

**Telephone and in-person respondents significantly different by chi-square test ($p < .05$)

Recall that a lack of telephone numbers kept 66% of the mail nonrespondents from ever having a chance to respond by phone.

4.3. Characteristics of respondents and nonrespondents

Table 3 is an important table, for it enables us to see how closely the returns matched the characteristics of the total population after each phase of data collection. Information for several demographic characteristics was obtained from administrative data for the entire sample, allowing us to make comparisons regarding age, gender, race/ethnicity, and primary language. In addition, administrative claims data were examined for the two years prior to data collection to identify the presence of a chronic condition. A set of diagnoses was compared with a list of ICD-9 codes considered to be indicative of the presence of significant health conditions in adults. Sample members with one of these diagnoses were considered to have a diagnosed chronic condition for the purposes of these analyses. The costs associated with care were also examined. The total cost of services covered by Massachusetts Medicaid during the prior two-year period was calculated. To adjust for differing lengths of tenure in the Medicaid program, the expenditures recorded in the

Table 3. Characteristics of respondents from administrative records: Respondents to each phase of the study compared to the total sample^a (%(n))

	Cumulative % of respondents			No interview	Total sample
	Mail	Mail and phone	Mail, phone, and field		
Age					
Less than 34	42**	45*	49	49	49
35 or older	58	55	51	51	51
Total	100% (497)	100% (650)	100% (936)	100% (464)	100% (1,400)
Gender					
Male	22	22	22	30*	25
Female	78	78	78	70	75
Total	100% (497)	100% (650)	100% (936)	100% (464)	100% (1,400)
Race/ethnicity					
White-not Hispanic	72*	73**	68	67	68
Black-not Hispanic	9**	9**	11	17*	13
Hispanic	16	15	17	11**	15
Other	4	3	4	5	4
Total	100% (485)	100% (633)	100% (913)	100% (453)	100% (1,366)
Primary language					
English	58	61	63	59	62
Spanish	9	8	8	4**	7
Other	33	32	29*	37*	32
Total	100% (497)	100% (650)	100% (936)	100% (464)	100% (1,400)
Diagnosed chronic condition					
No	69	72	75	77	75
Yes	31**	28	25	23	25
Total	100% (497)	100% (650)	100% (936)	100% (464)	100% (1,400)
Average monthly expenditures					
Lower 50%	46	48	52	54	53
Higher 50%	54**	52*	48	46	47
Total	100% (497)	100% (650)	100% (935)	100% (463)	100% (1,400)

^aData weighted to adjust for oversampling of adults in SSI or other special programs

*Significantly different from total sample by *t*-test ($p < .05$)

**Significantly different from total sample by *t*-test ($p < .01$)

24-month period were divided by the number of months the member had been enrolled, producing an average expenditure per month.

The characteristics of all those sampled were compared with the characteristics of those responding by cumulated phase of data collection. The characteristics of nonrespondents were also compared with those of the total sample.

Mail respondents differed from the total population on four of the six variables presented in Table 3, including both health status related variables. All four of these variables (age, race/ethnicity, diagnosed chronic condition, and average monthly expenditures) show a clear pattern: the combined returns from all three phases look much more like the entire sampled population than the results from only one or two waves of contact. When the phone interviews were added to the mail returns, the resulting sample still significantly underrepresented those under age 35 and African-Americans, and overrepresented whites and those with higher than average health care costs. When the in-person data were added, the only significant bias remaining was an underrepresentation of those who do not speak English or Spanish.

4.4. Relationship of protocol to responses

Table 4 presents the four key ratings from the survey by phase in which responses were obtained. The purpose of this analysis is to look at the effect of nonresponse on the key survey estimates. The rating of overall health care was significantly different by phase of data collection ($p = .018$). It can be seen from the lower mean score that people who did not respond until contacted in-person at their homes tended to be more critical of their health care than those who responded during the first two waves of contact. The key practical question, though, is how the inclusion of the data from the various phases affects the aggregate estimates.

In Table 5, the results from mail versus phone responses and mail plus telephone versus field responses were examined by item type. The substantive survey items were divided into five groups: items that utilized an always-to-never response task (primarily describing experiences with the health plan and provider interactions); items that asked about problems obtaining a service; the global ratings using the 0 to 10 rating scale; screening questions that asked about the need for or use of services using a yes-no response task; and questions that asked about the number of visits to doctors' offices and emergency rooms. Using t -tests for the global ratings and questions about the number of visits, and chi-square tests for the other items, responses to 11 of the 40 items (27%) were significantly different

Table 4. Mean global ratings on a scale of 0 to 10 by phase of contact^a (mean (n))

	p^b	Mail	Phone	Field
Rating of primary doctor	.369	8.37 (303)	8.64 (117)	8.32 (204)
Rating of specialist	.129	8.59 (213)	8.52 (61)	8.12 (129)
Rating of health care	.018	8.25 (320)	8.58 (111)	7.91 (214)
Rating of health plan	.207	8.23 (485)	8.49 (148)	8.10 (283)

^aData weighted to adjust for oversampling of adults in SSI or other special programs

^bOne-way analysis of variance used

Table 5. Whether or not answers for items with each response task differed* by phase of data collection^a

Item type	Total	Mail vs tel		(Mail + tel) vs field	
		Yes	No	Yes	No
Rating of services and providers:					
Always Never ^b	17	1	16	0	17
Big problem-Not a problem ^b	6	1	5	1	5
0 to 10 scale ^c	4	3	1	0	4
Need for or use of services:					
Yes-No ^b	11	5	6	2	9
Number of visits ^c	2	1	1	0	2
Total	40	11	29	3	37

**p* # .05^aData weighted to adjust for oversampling of adults in SSI or other special programs^bChi square tests used^c*T*-tests of means used

between the mail and phone phase. When the combined data from the mail and phone phases were compared with responses from the in-person phase of data collection, only three of the 40 items (7%) were significantly different.

5. Discussion

We learned a great deal about the problems of getting a good response rate in a population that some may think of as hard to survey. With an overall refusal rate of 6%, unwillingness to respond at all turned out to be a comparatively small part of the nonresponse problem. The 68% response rate and 82% cooperation rate achieved here indicate that Medicaid enrollees are willing to respond if they are approached in a way that works for them. Overall, the primary source of nonresponse was difficulty in locating sampled individuals.

The data also emphasize the interaction between responding and the amount of resources devoted to data collection, both in terms of the number of contacts with potential respondents and in offering different ways to respond. Motivation and salience of the study appear to be major factors in whether respondents will complete and return mail questionnaires. Those in poor health and those over age 35 were much more likely to return an instrument by mail. Mail strategies give potential respondents a chance to review the content of the survey and decide whether they want to answer. One can reasonably infer that those who need more health care services and those over age 35 are more likely than others to see a health survey as relevant to their own interests. Offering a Spanish version of the instrument to everyone also stimulated returns from Spanish speakers, counteracting the usual finding that this population does not respond well to mail questionnaires.

The telephone protocol was successful in reaching some people who did not respond by mail: younger respondents, those in good health, and particularly English speakers. Given little real resistance to responding, contact by phone enlisted cooperation from some of the

less motivated subgroups if we could reach them. The limits to the phone phase data stem mainly from the difficulty in finding correct phone numbers. The address information was often incorrect, and many people could not be located through directory assistance. Not surprisingly, people who do not speak English or Spanish as a first language almost never responded during the phone phase, even though Spanish-speaking interviewers were available. Unless contact information for the Medicaid population can be improved, the potential for telephone data collection to increase response rates may be limited. However, this was not our experience in a survey of a privately insured population (Fowler 1998). There, contacting mail nonrespondents by telephone was more successful in improving response rates as well as the representativeness of the sample.

The in-person phase, while the most expensive, also suffers least from either dependence on respondent motivation or availability of phone numbers. Indeed, 46% of the nonrespondents to mail and phone data collection efforts proved able and willing to respond to an in-person interviewer. Face-to-face administration proved to be a particularly good method to collect data from mail nonrespondents who were African-American or whose primary language was not English. Overall, 74% of these previous nonrespondents who could be found ended up responding to an in-person interviewer.

Raising response rates is desirable in itself because it increases the credibility of data. However, the most important result is that improving response rates also reduced nonresponse bias. In each of the areas that we could assess using administrative data, we can demonstrate that the final group of respondents looked more like the total sample than the group of respondents that would have resulted based on either mail or mail plus telephone returns alone. In most cases, the final group of respondents was virtually identical to the total sample in the ways that we could assess. Although in-person data collections have declined over the past two decades, the optimal strategy for Medicaid populations may be a mail phase followed by an in-person protocol. However, this may not always be an option because of cost or available staff.

Finally, one can ask how raising response rates affects results. In this case, the evidence is mixed. Looking at the four ratings that were among the key results of the survey, there was only one rating that was clearly significantly different by phase. However, this rating of health care is possibly the most important measure in the survey. The evidence is that adding the telephone responses to those obtained by mail had a significant effect on over a quarter of the estimates from the survey. However, further addition of the responses from in-person interviews, while increasing the credibility of the data and the representativeness of the sample, had minimal effects on the resulting data.

There is controversy now about how important it is to raise response rates. Many surveys are done that achieve poor response rates. Two recent analyses of telephone survey results found that efforts to enlist cooperation from more respondents did not affect the results to a significant degree (Keeter et al. 2000; Curtin et al. 2000). Siemiatycki and Campbell (1984) report few differences between early and late responders to a community health survey when the initial mode was either mail or telephone. However, the data presented here provide a clear example of how efforts to increase mail response rates make the resulting samples and those from a combined mail and telephone protocol much more like the study population. Each data collection step made the resulting sample more

representative in terms of demographics, health status, and use of services. Combined data collection techniques not only increased response rates, but also achieved better representation of young adults, people who used fewer health services, people who report being in better health, and ethnic minorities. It may be understandable why some of these groups, such as young healthy adults, 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.

One could ask why we do not recommend solving the problems of nonresponse by simply weighting the respondents. Adjusting sample characteristics to known characteristics of the population is limited by what information is available about the sample (which is usually very limited). Weighting is only valuable for those estimates that are related to those known population characteristics, and those relationships are almost always unknown and unknowable. Thus, even if weighting this sample would produce some improvements in some estimates, it would not address the general problem of nonresponse error that the article is describing.

As discussed above, these data provide a good example of how multiple modes of data collection can affect who responds to a survey. The fact that 27% of the key survey estimates from mail and telephone respondents were different indicates that there is enough variation among the people responding at each of those phases to affect survey responses. The cumulative responses of mail and phone respondents, however, are not very different from responses obtained during the field phase of data collection (see Table 5). Thus we also have an example of how improving the representativeness of the sample does not necessarily have much effect on the substantive estimates.

The central question is what the implications are for protocols to collect data about enrollees' experiences with their health plans. By general survey research standards, response rates below 60% or 70% are often considered problematic and likely to be biased. The data from the telephone and in-person phases show that those concerns are not unfounded. There are important differences among respondents to the three phases of data collection. On the grounds of credibility alone, having a more representative inclusion of people in better health, of young adults, and of ethnic minorities is valuable. Moreover, it seems most likely that when there are differences, the data from combining returns from multiple modes of data collection are more credible and more likely to accurately reflect patient experiences overall.

These results clearly will not generalize to all protocols, topics, and populations. However, perhaps they can help us move toward the development of better models of when and how the rate of response really does matter.

There are reasonable concerns about the effects of data collection mode on responses (Dillman 2000). There is ample evidence, however, that many questions produce comparable data across modes. Hochstim (1967), for example, reports over 1,000 comparisons between mail, telephone, and in-person survey estimates, with only 51 that were significantly different. With careful instrument design and testing, the differential effects of mode of survey administration may be decreased or eliminated, thereby allowing researchers to take advantage of the benefits that can be obtained from multi-mode data collection.

6. References

- American Association for Public Opinion Research (AAPOR) (2000). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Ann Arbor, MI.
- Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, 413–428.
- Dillman, D.A. (2000). *Mail and Internet Surveys: The Tailored Design Method* (Second Edition). New York: John Wiley.
- Farley, D. (RAND) and Jael, E. (Research Triangle Institute). (2000). Personal Communications.
- Fowler, F.J., Jr., Gallagher, P.M., Stringfellow, V.L., Zaslavsky, A.M., Thompson, J.W., and Cleary, P.D. (2002). Using Telephone Interviews to Reduce Nonresponse Bias in Mail Surveys of Health Plan Members. *Medical Care*, 40, 190–200.
- Fowler, F.J., Jr., Gallagher, P.M., and Nederand, S. (1999). Comparing Telephone and Mail Responses to the CAHPS[®] Survey Instrument. In CAHPS[®] Supplement. M. Weinberger (ed.). *Medical Care*, 37, MS41–MS49.
- Fowler, F.J., Jr. (1998). Nonresponse Bias in Mail Surveys of Health Plan Members. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 576–580.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley.
- Groves, R.M. and Lyberg, L.E. (1988). An Overview of Nonresponse Issues in Telephone Surveys. In *Telephone Survey Methodology*, R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (eds). New York: John Wiley.
- Groves, R.M., Singer, E., and Corning, A. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly*, 64, 299–308.
- Hochstim, J.R. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association*, 319, 976–989.
- Hox, J.J. and de Leeuw, E.D. (1994). A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys: Applying Multilevel Modeling to Meta-Analysis. *Quality and Quantity*, 28, 320–344.
- Keeter, S., Miller, C., Kohut, A., Groves, R.M., and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64, 125–148.
- Lasek, R.J., Barkley, W., Harper, D.L., and Rosenthal, G.E. (1997). An Evaluation of the Impact of Nonresponse Bias on Patient Satisfaction Surveys. *Medical Care*, 35, 646–652.
- McHorney, C.A., Kosinski, M., and Ware, J.E. (1994). Comparisons of the Costs and Quality of Norms for the SF-36 Health Survey Collected by Mail Versus Telephone Interview: Results from a National Survey. *Medical Care*, 32, 551–567.
- Siemiatycki, J. and Campbell, S. (1984). Nonresponse Bias and Early Versus All Responders in Mail and Telephone Surveys. *American Journal of Epidemiology*, 120, 291–301.

Teitler, J.O., Reichman, N.E., and Sprachman, S. (2003). Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population. *Public Opinion Quarterly*, 67, 126–138.

Weinberger, M. (ed.) (1999). CAHPS[®] Supplement. *Medical Care*, 37, MS1–MS116.

Received February 2000

Revised September 2004