

**NOTES FROM THE FIELD:
EXPERIMENTS IN INFLUENCING RESPONSE RATES FROM MEDICAID ENROLLEES**

**Patricia M. Gallagher, Floyd Jackson Fowler, Jr., Vickie L. Stringfellow, Center for Survey Research
P. M. Gallagher, CSR, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125**

Key Words: Survey methodology; Medicaid; Spanish; Mail; Telephone; In-person

Introduction

Developing methods to maximize response rates is a primary concern of survey researchers. The higher the percentage of returns from among those sampled, the more confidence in the generalizability of the findings. Certain segments of the population, for example, people with low household income, have demonstrated low rates of questionnaire return relative to the general population (Krysan, et al. 1994).

Medicaid programs throughout the United States are surveying members to evaluate the quality of their care experiences. These surveys are most often done by mail, sometimes with telephone follow-up of mail nonrespondents to increase response rates. Even when a combination of mail and telephone data collection strategies are used, the response rates are often disappointingly low.

To identify survey methods that maximize response rates in this population and to gain a better understanding of the significance of nonresponse, we conducted a series of studies among Medicaid enrollees. This paper presents the results of methodological experiments in the Medicaid population in Massachusetts to evaluate how response rates are affected by variations in: (1) instrument length; (2) the way that respondents are presented the opportunity to respond in Spanish; and (3) mode of administration.

There are many issues that must be addressed in designing a study: This paper examines three. First, there is a tension between the desire to collect a great deal of information from the population of interest and the need to limit respondent burden by creating parsimonious questionnaires. A second tension is the need to keep administrative costs down while making the survey accessible to Spanish-speaking members of the target population. If it is possible to use administrative records to identify those who might benefit from dual-language questionnaires, or to devise a method for respondents to request a Spanish instrument, survey costs can be minimized. Third, there is the question of whether additional efforts to reach a maximum numbers of respondents reduces nonresponse bias. Traditionally, three methods have been employed: self-administered mail questionnaires, interviewer-administered telephone interviews, and in-person

interviews. Face-to-face interviews usually involve higher costs per response than the other two methods (Hox & De Leeuw 1994), but generally demonstrate higher response rates than either mail or telephone interviews (Krysan et al. 1994).

Nonresponse bias occurs when nonrespondents differ systematically from respondents in ways that are relevant to what is being measured (Groves & Lyberg 1988). Previous studies have found that mail nonresponders are more likely to have less than a high school education, to be male, younger, non-white, and unmarried than mail respondents (Mc Horney, Kosinski, & Ware 1994; Lasek, et al. 1997).

Frequently, mail and telephone protocols are combined, and evaluative studies of these dual-phase methods are well documented (Dillman & Tarnai 1991). It is relatively rare, however, that the three basic data collection techniques are combined and nonresponse bias across three phases of administration in a single sample is evaluated. The findings from two early general population studies that did so (Hochstim 1967; Thornberry, 1976) suggest that using a variety of methods allows for the relative strengths of one protocol to compensate for the weaknesses of other methods.

Sample Design

The sampling frames for these tests were provided by the Division of Medical Assistance (DMA) for the 1997 (Survey A) & 1998 (Survey B) member surveys. A random sample of adults age 18 to 64 (n=5,747) and children age 17 or younger (n=8,322) was drawn for Survey A. Data collection followed a standard mail survey protocol.

In Survey B, we sampled 1600 adults and 1600 children and made sequential attempts to contact respondents -- first by mail, then by telephone, and finally, in-person all who had not responded either by mail or telephone. Half of each age group was randomly selected from members who receive SSI benefits, indicating the presence of a chronic medical condition, while the other half was a probability sample selected from the remainder of the MassHealth population.

Instrument Design

In Survey A, all sample members (adults and the parents of the selected children) were mailed a shortened (23 item) version of the 1997 NCQA HEDIS Member Satisfaction Survey.

Instrumentation for Survey B consisted of adult and child versions of CAHPS® 1.0 questionnaires. Within each sampled group, a random half received a shorter questionnaire (adults 54 items, children 65 items), while the rest received a longer instrument (adults 76 and children 94 items).

A critical feature of the instrument was that it was designed to produce comparable data whether it was interviewer- or self-administered. Questions were worded so they were identical in both forms. Testing suggested that very few of the questions are affected by mode of data collection (Fowler, Gallagher, & Nederland 1999).

Because a substantial portion of those receiving MassHealth benefits speak Spanish as their primary language, respondents had the option of completing the survey in either Spanish or English at each phase of the Survey B multi-mode data collection protocol. In the mail phase, half of the sample was sent an English questionnaire with an attached postcard that could be mailed back if the respondent preferred to fill out a Spanish version of the questionnaire. The other half was sent a single dual-language questionnaire printed in English on one side and Spanish on the other.

Data Collection Protocols

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 10 days later a reminder/thank you post card was sent. About two weeks after the initial mailing, a replacement questionnaire packet was sent to all those who had not yet responded. In Survey A, telephone reminder calls were made to non-responders about a month after the field period began, while Survey B involved the intensive follow-up of non-respondents described below.

Telephone Phase. Telephone protocols included a minimum of 6 call attempts to each case for which a telephone number could be identified. Calls were placed at a different times of the day and on different days of the week, with day time and evening call attempts occurring midweek and on weekends.

In-Person Phase: After exposure to complete mail and telephone protocols, all remaining non-responding cases were transferred to a field interviewer who attempted to contact the respondent at the last known address. The in-person interview protocols were similar to those outlined for the telephone; interviewers were required to make a minimum of 6 attempts, including Saturday and evening visits, spread out over at least a 2 week period. Extensive efforts were made to locate sample members. Interviewers tried to find those who did not live at the addresses originally provided by Medicaid (or updated through mail and

telephone efforts) by talking with neighbors and attempting to contact the landlord at the enrollee's last known address. Those contacted were usually interviewed in-person, but interviewers were allowed to offer the option of completing a self-administered form, which the interviewer would pick up. In-person interviewers were also permitted to conduct phone interviews with enrollees for whom they were able to obtain telephone numbers.

In the phone and in-person phases, bi-lingual interviewers were assigned to Spanish-speakers who were unable to complete the interview in English.

Analysis Plan

Instrument Length Test. Response rates for the three instruments of varying lengths were compared for both sampled groups, enrolled adults and parents of enrolled children. Only responses to the mail survey protocol are included in this analysis, i.e., all responses to Survey A and responses received during the mail phase of Survey B. All response rates reported in this paper were calculated as the proportion of eligible sample members responding, with sample members for whom we could not obtain good contact information considered eligible.

Instrument Language Tests. While Survey B involved 3 phases of contact, to mimic a dual mode data collection strategy, only data collected during the mail and telephone phases were used in these analyses. First, the language respondents chose to respond in was compared with type of instrument they had received. The goals were to learn how many respondents used the Spanish translation and which mode of delivery was the most effective in eliciting responses from Spanish speakers.

Second, in order to examine the feasibility of targeting Spanish language instruments to Spanish speaking respondents, the language the respondent chose was compared with an indicator of the enrollee's primary language from Medicaid administrative records.

Mode of Administration Test. The data collected from the adult samples in Survey B were analyzed to examine the procedural implications of the multi-phase approach to data collection, and to assess how nonresponse at each phase affected both representativeness and key descriptive results. To do this, we carried out four types of analyses: (1) Response rates, by phase of contact and cumulatively, were calculated for both subsamples (adults on SSI and those not enrolled in SSI); (2) Self-reported characteristics of respondents from each phase were compared (To adjust for the disproportionate sampling in the SSI group, in this and all subsequent analyses, weights were applied to adjust for differences in the probabilities of

selection.); (3) To provide information about those who never responded, administrative claims data were examined for the two years prior to data collection to identify the presence of a chronic condition and rates of utilization. The five most frequently occurring diagnoses in those claims were identified for each case. These diagnosis codes were compared with a list of ICD-9 codes considered to be indicative of the presence of a significant health condition in adults. The costs associated with care were also examined. The total cost of services covered by MassHealth during the prior two-year period was calculated. To adjust for differing lengths of time in the Medicaid program, the costs recorded in the 24-month period were divided by the number of months during that period the member had been enrolled in MassHealth, producing an average expenditure per month. Using that variable we divided covered enrollees into quartiles. To assess how each data collection phase affected representativeness, SSI enrollment, the average expenditure per month, and other demographic information available from administrative records (language, race, age, and gender) were compared for respondents at each phase of contact, those who never responded, and the total sample; (4) To determine if improving response rates by offering respondents multiple ways to respond affects key survey results, respondents' global ratings of their primary care providers, specialists, overall health care, behavioral health care, and health plans were compared across the three phases of administration.

RESULTS

Instrument Length Test Results. The number of questions that respondents were asked to answer, from as few as 23 to as many as 95 items, had little effect on rate of return of the instruments. Response rates by mail ranged from 34-38%. There were no significant differences in the response rates obtained using questionnaires of different length.

Instrument Language Test Results. The specific effect of the dual-language instrument is to increase the number of people responding in Spanish. The proportion of respondents returning a Spanish language questionnaire increased from 5% of those who received an English instrument with a postcard to request the Spanish version to 17.5% of those who received the Canadian-style instrument.

Just 44 of the original sample of 1600 returned a postcard requesting a Spanish-language instrument, but 82% of these motivated respondents completed the questionnaire.

Employing either a dual-language questionnaire or a postcard request protocol eliminates reliance on Medicaid records for targeting Latino households to

receive a Spanish-language instrument. Analysis of the language in which respondents chose to respond implies that the administrative identification of Spanish speakers is imperfect; some respondents identified as non-Spanish speakers elected to respond in Spanish, while some Spanish speakers returned an English questionnaire. More than half of those who responded in Spanish were not identified as Spanish speakers in Medicaid records.

Mode of Administration Test Results.

Response Rates by Phase of Contact. Table 1 presents cumulative response rates for each sample group by stage of respondent contact. Predictably, each successive contact improved response rates, but the telephone protocol was the least productive of the three phases employed. This table shows that those with no known chronic conditions were more likely to respond than those enrolled in SSI (73% versus 63%).

For nearly 15% of the total sample the contact information was incorrect. When the sample is stratified by eligibility for SSI, it can be seen that it is much more difficult to locate members in these programs. We were not able to trace 18% of the sample with a known chronic condition compared with 12% of those without. This difference may reflect infrequent updating of administrative records and the relatively longer periods of Medicaid eligibility that enrollees with chronic conditions have compared to those who are income eligible.

Another way to think about outcome rates is to calculate the rate of cooperation. This allows an estimation of respondents' willingness to participate that is independent of the quality of contact information provided. This is the proportion of all eligible units ever contacted who responded. The overall cooperation rate for the three protocols was over 80%.

Overall, about 34% of the eligible sample responded by mail, another 10-13% were picked up by telephone, and the in-person effort brought the cumulative response rate to 68%. Refusals did not prove to be much of a problem, with only about a 6% refusal rate overall.

Respondent Characteristics. Without an interviewer to encourage compliance, the relevancy of the topic can be particularly important to mail responders. Respondents who self-report being in fair or poor health were more likely than those in better health to respond by mail, as were sampled individuals who were 35 or older. The salience of health-related matters may be a factor in these findings.

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 nonresponders

tended to increase responses from younger people, those with higher levels of education, and among whites. At the same time, those who reported speaking a language other than English at home tended not to respond over the telephone, but were relatively willing respondents by mail and in-person.

Table 2 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. The characteristics of all those sampled are in the right-hand column. Of the six variables presented, five show a clear pattern: returns from all three phases look much more like the entire sampled population than the results from only one or two waves of contact. Because health is the primary focus of this survey, it is particularly important to note how closely the results of the combination of mail, phone, and in-person efforts mirror the sample with regard to chronic conditions and expenditures per month. The only nonresponse bias from the mail returns that did not improve in subsequent phases was the under representation of males.

Relationship of Protocol to Responses. People who did not respond until being contacted in-person at their homes tended to be more critical of their health care than mail and telephone respondents. The rating of overall health care was significantly different by phase of data collection ($p=.02$). The other ratings did not differ significantly across phases.

When responses were examined by phase for three types of items -- screening questions about need or use of services, items with an always-to-never response category (primarily describing experience with the health plan and provider interactions), and items that asked about problems obtaining a service -- over two thirds of the questions demonstrated no differences. Only one of the Always-to-Never and one Big Problem-to-No Problem item were significantly different by phase, but 6 of the 8 screening (Yes-No) questions were significantly different by phase. The CAHPS® reporting protocol involves 5 multi-item composites; only one of the five composites, the difficulty in handling plan paperwork, was significantly different by phase of data collection. However, 6 of the 8 screening items about need or use of assorted services varied significantly by phase of data collection. In general, those who used more services were more likely than average to respond to the mail protocol.

DISCUSSION

The biggest source of nonresponse was difficulty in locating the sampled individuals -- unwillingness to respond turned out to be a very minor factor. Medicaid enrollees will respond if they are approached in a way that works for them. In the three-mode protocol, we

obtained a cooperation rate of over 80% from among those who had any chance at all of responding.

Of course, factors other than instrument length, language, and mode of survey administration can affect response rates. These include the survey topic, item content, quality of respondent contact information, and number of respondent contacts. A complete discussion of these factors is beyond the scope of this paper. Moreover, none of the response rates achieved by mail alone in this study meet desirable standards for survey returns. Using a combination of mail, telephone, and face-to-face approaches, response rates of 63% for children and 73% for adults were achieved in this population of Medicaid enrollees.

CONCLUSIONS

Medicaid enrollees were about as likely to complete a relatively long questionnaire as a shorter one. The results of the instrument length test also demonstrate that a short questionnaire alone is not enough to achieve high response rate in this population.

The additional printing and mailing expenses associated with a dual-language Canadian-style questionnaire are justified by an increase in Spanish language returns from 5% to 13%. At the same time, there was no adverse affect on the response rate of English speakers associated with use of the two-language questionnaire. If fielding a dual-language instrument is not feasible, inviting respondents to use a postcard request will help in getting responses from Spanish speakers. However, because the record-based method of primary language identification was not perfect, targeting dual-language instruments to those identified as Spanish speakers, in this case, would have missed more than half of those who responded in Spanish.

The data also emphasize the interaction between responding and the mode of data collection. Motivation and salience of the study appear to be major factors in completing and returning a mail questionnaire. Those in poor health and those over 35 were much more likely to return an instrument by mail; one can reasonably infer these groups are more likely than others to see a health survey as relevant to their own interests.

The telephone protocol was successful in reaching some people who did not respond by mail: younger respondents, those in good health, and particularly English-speaking whites. Given little real resistance to responding, the phone enlisted cooperation from some of the less motivated subgroups if we could reach them. The biases in 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 in the phone phase.

Unless contact information for the Medicaid population can be improved, telephone may be an inadequate way to survey this population. This was not our experience, however, in a privately insured population (Fowler 1998). There telephone contact was much more successful in improving the representativeness of a sample first contacted by mail.

The in-person phase, while the most expensive, also suffers least from either dependence on respondent motivation or availability of phone numbers, and, indeed, 25% of the sample who could not or would not respond by mail or phone proved able and willing to respond to an in-person interviewer. It was a particularly good way to collect data from mail nonrespondents who were African-American or whose first language was not English.

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 based on administrative data, we can demonstrate that the final sample looked more like the total population than the samples that would have resulted based on mail or combined mail and telephone returns. In most cases, the final sample was virtually identical to the total population 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.

Finally, one can ask how raising response rates affects results. In this case, the evidence is mixed. Looking at the five ratings that were among the key results of the survey, there was only one item that was clearly significantly different by phase. However, this rating of health care is arguably the most important measure in the survey. On the other hand, only two of the 17 reports of experiences with getting health care differed by phase of data collection.

There is controversy now about how important it is to raise response rates. Many surveys achieve poor response, and some analyses find that efforts to enlist cooperation from more respondents do not affect the results to a significant degree. These data provide a clear example of how efforts to increase response rates make the resulting samples much more like the study population. The data also provide a good example of how mode of data collection can affect who responds to a survey. The effect of increased representativeness on key survey estimates was mixed but at least a few key results were different, and better, when the response

rate was raised.

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.

REFERENCES

- Dillman, DA, Tarnai J. 1991 "Mode effects of Cognitively Designed recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In *Measurement Errors in Surveys*, PN Biemer et al (editors), pp. 367-93. NY: Wiley.
- Fowler, FJ, Jr., Gallagher, PM, & Nelderland, S. (1999) Comparing telephone and mail responses to the CAHPS survey instrument. In Weinberger, M. (Ed). CAHPS® Supplement. *Medical Care*, 37 (3): Pp. MS41-MS49.
- Groves, RM & Lyberg, LE (1988). An overview of nonresponse issues in telephone surveys, pp 191-212 in RM Groves, PP Biemer, LE Lyberg, JT Massey, WL Nicholls II, & J Waksberg (eds.) *Telephone Survey Methodology*. NY:Wiley
- Hochstim, JR (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 62 (319), 976-989.
- Hox, JJ & De Leeuw, ED. (1994). A comparison of nonresponse in mail, telephone, and face-to-face surveys: Applying multilevel modeling to meta-analysis. *Quality & Quantity*, 28: 320-344.
- Krysan, M, Schuman, H, Scott LJ, & Beatty, P (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly*, 58:381-399.
- Lasek, R.J., Barkley, W., Harper, D.L. & Rosenthal, G.E. (1997). An evaluation of the impact of nonresponse bias on patient satisfaction surveys. *Medical Care*, 35(6), 646-652.
- McHorney, C.A., Kosinski, M., & 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, in *Medical Care*, 32(6), 551-67.
- Perkins, JJ & Sanson-Fisher, RW (1998). An examination of self- and telephone-administered modes of administration for the Australian SF-36. *Journal of Clinical Epidemiology*, 51 (11): 969-973.
- Thornberry, OT (1976). *An Evaluation of Three Strategies for the Collection of Data from Households*. Doctoral dissertation, Department of Sociology, Brown University.

TABLE 1. CUMULATIVE RESPONSE RATE FOR ADULTS AT EACH PHASE OF DATA COLLECTION

	Ineligible	Eligible Sample*	Refusal	Never Located	Completed Survey	Percent Responding	Cumulative Response Rate
Non-SSI Adults							
Mail	1	799	1	48	273	34%	34%
Phone	13	786	26	57	99	13%	47%
Field	16	770	19	92	192	25%	73%
Total	30	770	46	92	564	73%	--
SSI Adults							
Mail	1	799	6	144	299	37%	37%
Phone	8	791	32	52	72	9%	47%
Field	9	782	12	140	125	16%	63%
Total	18	782	50	140	496	63%	--
Overall Total	48	1552	96	232	1060	--	68%

* Eligible sample adjusted to reflect new information obtained at each wave.

Table 2: Characteristics of Respondents Identified through Administrative Records at Each Phase of the Study (Weighted Data)

	Cumulative Percent of Respondents to Each Phase of Study				
	Mail (%)	Mail & Tel (%)	Mail/Tel/Field (%)	No Interview (%)	Tot. Sample (%)
Age					
less than 34	43	45	49	49	49
35 or older	57	55	51	51	51
Total	100 (n=497)	100 (n=153)	100 (n=286)	100 (n=465)	100 (n=1401)
Gender					
Male	22	22	22	30	25
Female	78	78	78	70	75
Total	100 (n=498)	100 (n=153)	100 (n=286)	100 (n=465)	100 (n=1403)
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 (n=485)	100 (n=149)	100 (n=280)	100 (n=453)	100 (n=1367)
Primary Language					
English	58	61	63	59	62
Spanish	9	8	8	4	7
Other	33	31	29	37	31
Total	100 (n=498)	100 (n=153)	100 (n=286)	100 (n=464)	100 (n=1401)
Chronic Condition by Dx					
No	77	78	80	83	80
Yes	23	22	20	17	20
Total	100 (n=806)	100 (n=249)	100 (n=505)	100 (n=696)	100 (n=2256)
Avg Expenditures/Month					
\$190.32 and up	27	26	25	26	25
\$88.19 to \$190.31	31	29	27	24	26
\$49.41 to \$190.30	21	23	25	22	24
\$49.40 or less	21	23	24	28	25
Total	100 (n=496)	100 (n=154)	100 (n=287)	100 (n=464)	100 (n=1401)