

## Evaluating Mode Effects in the Medicare CAHPS Fee-For-Service Survey

Norma Pugh, MS, Vincent Iannacchione, MS, Trang Lance, MPH, Linda Dimitropoulos, PhD  
RTI International, Research Triangle Park, NC 27709

**Key Words:** Mode of Data Collection, Social Desirability, CAHPS, Non-Response Bias

### Introduction

This research seeks to determine how the mode of non-response follow-up (specifically, overnight delivery or telephone administration) to a mail survey of beneficiaries enrolled in Medicare Fee-For-Service (MFFS) (a.k.a. Original Medicare) affects survey estimates of CAHPS (Consumer Assessment of Health Plans Study) performance measures. The CAHPS-MFFS Survey is one of three CAHPS surveys of Medicare beneficiaries conducted annually by the Centers for Medicare & Medicaid Services (CMS). The CAHPS surveys fulfill a requirement of the Balanced Budget Act of 1997 to provide information to Medicare beneficiaries on the quality of health services provided through the Original Medicare Plan and to compare this information to similar information collected from beneficiaries enrolled in Medicare managed-care health plans.

For the 2000 CAHPS-MFFS Survey, we selected a stratified random sample of 168,000 beneficiaries from more than 30 million Medicare FFS beneficiaries residing in the U.S. and Puerto Rico in August, 2000. We achieved a 54.3% response rate after two mailings of the survey instrument. Beneficiaries who did not respond to either mailing were placed into telephone follow-up if a telephone number could be obtained for them. Beneficiaries for whom a telephone number could not be obtained were sent an overnight delivery of the survey instrument if they resided in a county with low response to the mail portion of the survey. With the non-response follow-up, we achieved a 63.9% response rate among beneficiaries who were eligible for the survey.

For this research, we studied non-respondents to the mail portion of the 2000 MFFS Survey who resided in the 53 (primarily urban) counties of the U.S. where response rates were lowest. These counties accounted for 23.8% of all Medicare FFS beneficiaries in 2000. Among the 7,773 beneficiaries who were followed-up via overnight delivery in these counties, 1,168 (15.0%) responded. Among the

9,817 beneficiaries who were followed-up via telephone in the same counties, 1,599 (16.3%) responded.

As **Table 1** shows, we did not find significant mode effects associated with a beneficiary's overall rating of health care or with any of the 12 report items that comprise the three CAHPS composites related to communication with physician or nurse, getting needed care, and getting care quickly. However, we did find that 56.1% of beneficiaries who were followed-up by telephone gave Medicare the highest rating possible compared to only 44.0% of beneficiaries who were followed-up by overnight delivery. In addition, we found that only 39.7% of beneficiaries who were followed-up by phone reported an overall health status of fair or poor compared to 48.5% of overnight respondents. We observed a similar but smaller difference between telephone respondents reporting fair or poor mental health (26.8%) and overnight respondents (32.1%). In the remainder of the paper, we develop three possible explanations for these apparent mode effects.

### Three Possible Explanations for Apparent Mode Effects

**Explanation 1:** *Mode effects are the result of a preference by certain types of persons to respond to the follow-up by telephone rather than by overnight delivery.* For example, persons with little or no formal education may have been more likely to respond to a telephone interview than to a self-administered questionnaire. Similarly, older persons with writing difficulties caused by infirmities such as arthritis may have preferred to be interviewed by telephone. These preferences could explain the differences in overall and mental health status between telephone and overnight respondents as well as differences in their ratings of Medicare.

The assessment of whether certain types of persons were more willing to respond by phone than by overnight delivery was confounded by the fact that persons selected for the follow-up survey were not randomly assigned to a mode of follow-up. Instead, all persons for whom a telephone number could be found were placed into the telephone portion of

**Table 1. CAHPS Questions Examined for Mode Effects<sup>1</sup>**

CAHPS Ratings	<b>How would you rate your experience with Medicare?</b> How would you rate your health care from all doctors and other health providers?
Health Ratings	<b>In general, how would you rate your overall health now?</b> <b>In general, how would you rate your overall mental health now?</b>
Questions Comprising the CAHPS Getting Needed Care Composite	How much of a problem, if any, was it to get a personal doctor or nurse you are happy with? How much of a problem, if any, was it to see a specialist that you needed to see?  How much of a problem, if any, was it to get the care you or a doctor believed necessary? How much of a problem, if any, were delays in health care while you waited for approval from Medicare?
Questions Comprising the CAHPS Communication Composite	How often did doctors or other health providers listen carefully to you?  How often did doctors or other health providers explain things in a way you could understand?  How often did doctors or other health providers show respect for what you had to say?  How often did doctors or other health providers spend enough time with you?
Questions Comprising the CAHPS Getting Care Quickly Composite	When you called during regular office hours, how often did you get the help or advice you needed?  How often did you get an appointment for regular or routine health care as soon as you wanted?  When you needed care right away for an illness or injury, how often did you get care as soon as you wanted?  How often did you wait in the doctor's office or clinic more than 15 minutes past your appointment time to see the person you went to see?
<sup>1</sup> Questions in bold text exhibited apparent mode effects. Except for health status, all questions refer to a six-month reference period.	

the follow-up. All others were placed into the overnight-delivery portion of the follow-up. Because of the lack of randomization to mode of follow-up, the composition of persons assigned to each mode of follow-up was influenced by our ability to find telephone numbers for as many follow-up sample members as possible.

**Table 2** illustrates the differences in the composition of sample members in the telephone follow-up and those in the overnight follow-up. For example, beneficiaries who lived in California made up 27% of the overnight delivery group and only 16% of the telephone group. This disparity may be caused by the high

percentage of unlisted telephone numbers among California residents. Unlisted telephone numbers are more difficult to obtain than listed numbers. There were no significant differences in the sample composition with respect to utilization measures derived from Medicare claims data.

We tested this possible explanation of mode effects by developing a logistic regression model using response to the follow-up (yes or no) as the outcome, mode of follow-up as a predictor of response, and demographic and health-care utilization measures as covariates.

**Table 2. Demographic Composition of the Follow-Up Sample**

	% of Overnight Sample	% of Telephone Sample
Under 65 years of age	21%	13%
Non-white	35%	25%
Dual Medicaid Eligibility	33%	18%
California Resident	27%	16%

**Table 3. Significant Predictors of Response to Follow-up**

Demographic Predictors		Utilization Predictors	
Previous managed care	(+)	One or more in-patient claim	(+)
Representative payee	(-)	One or more out-patient claim	(+)
Age	(-)	One or more SNF claim	(-)

(+) denotes an increased propensity to respond.

(-) denotes a decreased propensity to respond.

**Table 3** summarizes the significant predictors of response to follow-up. The model indicated that mode was not significantly associated with likelihood of response after adjusting for covariates such as age, the occurrence of inpatient or outpatient stays, and the use of a skilled nursing facility. We noted an increased propensity for response among beneficiaries who were previously in managed care; who had at least one in-patient claim; or who had at least one out-patient claim. There was a decreased propensity for response among beneficiaries who had a representative payee (agent), were older, or had at least one skilled nursing facility (SNF) claim.

**Explanation 2: Mode effects are the result of demographic differences between telephone and overnight respondents.** To test this hypothesis, we developed three logistic regression models: one for giving Medicare the highest possible rating, one for an overall health status of fair or poor, and one for a mental health status of fair or poor. In each of these models we included mode of follow-up as a predictor along with the demographic and utilization covariates shown in **Table 4**. We used segmentation analysis to identify the interactions that were included in each model. Individual terms then were removed, one or two at a time, based on their significance level.

**Factors Affecting the Rating of Medicare.** A segmentation of follow-up respondents

suggested a number of interactions. For example, education was the largest discriminator for this sample, with those having less than a high school education rating Medicare 15.0 percentage points higher than those with a high school education. Further, among follow-up respondents with less than a high school education, the presence or absence of a proxy resulted in a 17.3 percentage point difference in the Medicare rating. Among follow-up respondents with a high school education, mode of response was the largest discriminator, with a 17.7 percentage point difference.

**Factors Affecting Poor/Fair Overall Health Status.** As in the ‘Rate Medicare a 10’ model, mode of response and several demographic predictors were significant in the final ‘Poor/Fair Health Status’ model. The segmentation showed dual Medicare/Medicaid eligibility as the largest discriminator, with those who were dually eligible reporting poor/fair health status at a rate of 21.2 percentage points higher than those who were not dually eligible.

**Factors Affecting Poor/Fair Mental Health Status.** The segmentation of follow-up respondents showed that presence/absence of a proxy was the largest discriminator for the ‘Poor/Fair Mental Health Status’ sample group. Beneficiaries who responded by proxy reported poor/fair mental health status at a rate of 23 percentage points higher than those who did not respond by proxy.

**Table 4. Covariates Used in the Logistic Regression Models<sup>1</sup>**

Candidate Covariates	Highest Rating of Medicare	Poor/Fair Overall Health	Poor/Fair Mental Health
<b>Demographic Covariates</b>			
California Resident			
Last Year of Life	X	X	X
Education	X	X	X
Dual Eligibility	X	X	X
Representative Payee		X	X
Gender	X		
Race	X		
Previous Managed Care	X		
Proxy	X		X
Age	X		X
Poor/Fair Overall Health	X	--	X
Poor/Fair Mental Health	X	X	--
<b>Utilization Covariates</b>			
One or More In-Patient claims		X	
One or More Out-Patient claims	X	X	
One or More Home Health Claims		X	
One or More SNF Claims			

<sup>1</sup>X identifies covariates that were significant at the 0.05 level.

We computed predictive margins (Korn and Graubard 1997) to estimate the effect of mode on each of the three outcome measures ('Highest Rating of Medicare,' 'Poor/Fair Health Status', and 'Poor/Fair Mental Health Status') after adjusting for significant covariates. The predictive margins shown in **Table 5** may be viewed as the expected outcome if the entire follow-up had been administered by overnight delivery or by telephone interview. For example, the observed percent of telephone respondents who rated Medicare highest is 12.1 percentage points higher than the rating for overnight respondents. However, we estimate that even if

the follow-up sample had been implemented entirely by telephone, the percent rating Medicare highest would still be 9.1 percentage points higher than a follow-up sample implemented entirely by overnight delivery. While the inclusion of demographic factors did explain a portion of the differences between the answers given by telephone and overnight respondents, the mode effect remained significant in all three models and led us to speculate whether the telephone follow-up induces social desirability.

**Table 5. Influence of Follow-Up Mode on Outcome Measures**

Mode of Follow-Up	Highest Medicare Rating <sup>1</sup>		Poor/Fair Overall Health <sup>2</sup>		Poor/Fair Mental Health <sup>3</sup>	
	Observed Rate % (SE)	Predictive Margin % (SE)	Observed Rate % (SE)	Predictive Margin % (SE)	Observed Rate % (SE)	Predictive Margin % (SE)
<b>Overnight Delivery</b>	44.0 (0.02)	44.3 (0.02)	48.5 (0.01)	50.1 (0.02)	32.1 (0.01)	32.7 (0.01)
<b>Telephone</b>	56.1 (0.01)	53.4 (0.02)	39.7 (0.01)	42.3 (0.01)	26.8 (0.01)	27.5 (0.02)

<sup>1</sup> p-value for predictive margin (overnight delivery versus telephone) < 0.001

<sup>2</sup> p-value for predictive margin (overnight delivery versus telephone) < 0.001

<sup>3</sup> p-value for predictive margin (overnight delivery versus telephone) = 0.01

**Explanation 3: Mode effects are the result of “social desirability” among telephone respondents.** Social desirability is the tendency for a person to respond in a manner that is believed to be socially acceptable to the person making the inquiry. Respondents share a tendency to want to please the interviewer by responding in a way that they believe to be consistent with the beliefs held by the interviewer.

Mode effects between telephone and mail survey responses are frequently observed and can be quite pronounced (Dillman, 2000). The most consistent finding in the literature is that social desirability differences are frequently found in data collected from interviews than in data collected from self-administered surveys (DeMaio, 1984; Dillman, Sangster, Tarnai, and Rockwood, 1996). This is especially true when asking potentially sensitive or embarrassing questions (Do you ever drink and drive?) but it is also true when asking questions that are considered fairly innocuous (How would you describe your current health?). The social desirability literature consistently shows that when asked to rate their general health, a greater proportion of respondents will choose the more negative response choices (fair, poor) when the survey is self-administered than when the questions are asked by an interviewer. Similarly, if the respondent believes that the interviewer represents the Medicare program and that the interviewer holds positive beliefs about Medicare, then out of a desire to please the interviewer, the respondent, when asked to rate Medicare, can be expected to answer using the more positive end of the scale.

### Summary and Discussion

Most CAHPS measures were not affected by mode of follow-up to the 2000 CAHPS MFFS Survey. However, the prevalence of the highest rating of Medicare was 12.1 percent higher among beneficiaries followed up by telephone than those followed up by overnight delivery. In addition, the prevalence of self-reported poor/fair overall and poor/fair mental health was 8.8 percent and 5.8 percent lower among those followed-up by telephone.

We developed a series of regression models to determine whether these apparent mode effects could be explained by either differences in the response propensities of persons selected for the telephone and overnight

follow-up samples or by differences in the demographic and health utilization measures of those who responded to the follow-up. Our models indicated that mode did not significantly affect a person’s propensity to respond to the follow-up, and that mode remained a significant factor for these differences even after adjusting for demographic differences between telephone and overnight respondents. These results led us to speculate that mode may induce social desirability (i.e., instinctive desire to please on telephone) in the follow-up sample estimates.

Our current plan is to embed a randomized mode effects experiment into the 2002 CAHPS MFFS Survey so that we can get a better idea of the influence of mode on the follow-up sample estimates. If we find more conclusive evidence of mode effects, we may consider including mode as a case-mix adjustment variable for CAHPS measures.

### References

- DeMaio, T.J. (1984). Social desirability and survey measurement: A review. C.F. Turner & E. Martin (Eds.), *Surveying subjective phenomena: Volume 2* (pp.257-282).
- Dillman, D.A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. 2<sup>nd</sup> Ed. Wiley: New York, (pp. 224-227).
- Dillman, D.A., Sangster, R.L., Tarnai, J., & Rockwood, T. (1996). Understanding differences in people’s answers to telephone and mail surveys. M.T. Braverman & J.K. Slater (Eds.), *New directions for evaluation series*, 70 (Advances in survey research), (pp. 45-62. San Francisco: Jossey-Bass.
- Korn EL, Graubard BI (1997). Predictive Margins with Survey Data. *Proceedings of the American Statistical Association, Survey Research Methods Section*, 651-656.
- RTI and RAND (2001). *Implementation of Medicare CAHPS Fee-For-Service Survey Final Report for Year 1*, Prepared for the Centers for Medicare & Medicaid Services, Contract No. 500-95-0061.