

A Modeling Approach: Use of Survey Contact Day and Time to Model Response Propensity in the Medical Expenditure Panel Survey

Karen E. Davis¹

¹Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

Abstract

The Medical Expenditure Panel Survey (MEPS) is a nationally representative health survey that is used to produce estimates of health care expenditures and utilization. The MEPS uses the National Health Interview Survey (NHIS) as its sampling frame, allowing for limited paradata that may assist with developing discrete hazard models to assess the likelihood of a completed interview with sample households. Using data from the 2004 NHIS, along with 2005 MEPS contact history data, the purpose of this paper is to explore factors that may influence a successful interview in the MEPS.

Key Words: Sample survey, paradata, MEPS, contact history data

1. Introduction¹

The Medical Expenditure Panel Survey (MEPS) sponsored by the Agency for Healthcare Research and Quality (AHRQ) collects detailed information regarding the use and payment for health care services from a nationally representative sample of Americans. The household sample (MEPS-HC) is selected from participants in the previous year's National Health Interview Survey (NHIS), and contains about 13,000 families and 30,000 individuals. Although some data elements are similar in MEPS and NHIS, all MEPS data are collected independently and there is no formal reconciliation between the 2 surveys (Cohen, et. al, 2009). The MEPS-HC is designed to provide annual national estimates of the health care use, medical expenditures, sources of payment, and insurance coverage for the US civilian noninstitutionalized population. In addition to collecting data designed to yield annual estimates for a variety of measures related to health care coverage, use, and expenditures, the MEPS-HC also provides data on individuals' health status, demographic characteristics, employment, and access to health care, which can be aggregated to provide estimates for families and population subgroups of interest.

The MEPS-HC survey uses the NHIS as its sampling frame, and collects data through an overlapping panel design in which any given sample panel is interviewed a total of 5 times in person over a fielding period of about 30 months. The rounds of interviewing are spaced about 5 months apart. Therefore, MEPS full calendar year files contain data from the first year of a new panel combined with that of the second year of the previous panel. The MEPS interview is administered through a computer-assisted personal interview loaded onto a laptop computer, and takes place with a knowledgeable family respondent who reports for him or herself and for other family members.

2. Background

Data on the survey process, known as paradata, includes call records generated electronically or by an interviewer, observations of interviewers and respondents, audio recordings of interviewer and

¹ The findings and conclusions in this paper are those of the author and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

respondent interactions, as well as items generated by computer-assisted instruments, such as response times and key strokes (Kreuter and Casas-Cordero, 2010). Paradata can be used to measure survey quality, as well as to monitor survey fieldwork. Obtaining an interview requires the interviewer to successfully contact the sampled person. However, the steady decline in MEPS household response rates over the past 10 years (from 66 percent in 1999 to 57.2 percent in 2009) is consistent with increased noncontacts and refusals. Tables 1 and 2 highlight two important challenges for the first round of the MEPS, that is, lower response rate (79 percent versus 94.1 percent) and higher refusal rate (16.6 percent versus 4.5 percent). However in viewing these estimates, note that the Round 2 rates are conditional upon successful response to Round 1.

MEPS field interviewers typically call the sampled household to set up an appointment for the in-person visit. The interviewer records contact attempts for all sample households, regardless of whether contact is made. In addition to date, time and mode of attempt (phone versus in-person), interviewers report the outcome disposition of the contact (i.e., appointment, completed interview, noncontact). For attempts resulting in a noncontact, interviewers report why an interview did not occur upon making contact (e.g., not home, temporary refusal, callback, language problem). This information is collected in both the MEPS Record of Calls (ROC) and Paradata restricted-use files. Likewise, the NHIS has annual Paradata files to aid survey researchers in analyses about the data collection process. Because the MEPS uses the NHIS as its frame, not only is sampling information provided, but also contact timing data are available about the NHIS final interview. During training, MEPS interviewers receive a “face” sheet that provides the day and time of the final NHIS interview to aid in contacting sampled households for their first round interview.

This paper examines the timing of MEPS contact attempts and explores its effects on the propensity for a sampled case to respond to the household interview. For this paper, we explored the following two research questions:

Question 1: Are MEPS contact attempts on a similar day and time as the final NHIS interview more likely to result in higher Round 1 response and fewer numbers of contacts?

Table A illustrates the four scenarios for similar day and time of contact attempt that were used in this research. Morning and afternoon hours were combined into one category (8:00 a.m. through 5:50 p.m.), while evening hours began at 6:00 p.m. Weekday was defined as Monday through Thursday.

Table A. Four Scenarios for Similar Day and Time

	Weekday	Weekend
Morning/afternoon	1	2
Evening	3	4

Question 2: Using available MEPS paradata, what are some of the predictors that affect the propensity of a case to be interviewed on the next call attempt?

3. Methods

Using data from the MEPS and NHIS, we examined the contact attempts during 2005. The 2005 MEPS dataset is formed by combining Rounds 1, 2, and 3 of Panel 10 (which uses the 2004 NHIS sampling frame) with Rounds 3, 4, and 5 of Panel 9 (which uses the 2003 NHIS sampling frame). Analyses were restricted to the first three rounds of Panel 10 since the 2003 NHIS frame data were not available at the

time of this research. The first step in this process was to link the restricted-use 2005 MEPS ROC and Paradata files to the 2004 NHIS public-use data (NHIS, 2004). The 2005 MEPS Record of Calls file contains observations on each contact attempt with a sample unit, and for Panel 10 (Rounds 1, 2, and 3), this file contained 115,569 contact attempts. The 2005 MEPS Paradata file contains observations on each attempt yielding a contact, and includes variables such as the race/ethnicity of the respondent, the number of true contacts (defined as a count of all records of calls except those not home, moved, unable to enter structure, or unable to locate), whether the respondent is reluctant to respond, and whether the case was ever traced. For Panel 10 (Rounds 1, 2, and 3), the Paradata file contained 23,730 persons. The 2005 MEPS-2004 NHIS link file was used to facilitate linkage of the MEPS data files to the NHIS data. This file contains 33,961 person-level records, of these, 15,459 persons are linked to the 2004 NHIS. Records that could not be linked include newborns, newly in-scope persons as well as a small number of cases where the NHIS identified a household as responding but was actually a non-responding household. The final step in this process was to link these data to the 2004 NHIS restricted-use data file that contained variables related to the NHIS field work, such as the date and time of the final outcome, as well as information on the best time to contact the household. This file contained 47,611 persons with 52.3 percent (24,918) completed or sufficient partial interviews. After linking all of these files, a combined dataset was created containing 101,913 contact attempts for 7,198 persons. The vast majority, 70.2 percent, of these cases were first contacted by phone; while 29.6 percent were first contacted in person (0.2 percent was missing).

To estimate the propensity of a case to be interviewed on the next call, discrete hazard models using selected variables from the combined dataset were developed (Allison, 2010). Each individual's call history can be broken down into a set of discrete time units that are treated as distinct observations. After pooling these observations, the next step is to estimate a binary regression model predicting whether a contact did or did not occur in each time unit. Using Cox's proportional odds model for discrete-time data, we let P_{it} be the conditional probability that individual i has an event at time t , given that an event has not already occurred to that individual. The model says that P_{it} is related to the covariates by a logistic regression equation:

$$\log \left[\frac{P_{it}}{1 - P_{it}} \right] = \alpha_t + \beta_1 x_{it1} + \dots + \beta_k x_{itk}$$

Where $t=1,2,3,\dots$. This model is most attractive when events can only occur at regular, discrete points in time. We estimate the model by maximum likelihood, so that we get explicit estimates of the α_t s. This method gives us estimates for the effect of time on the odds of the event, as reflected in the α_t s. Because time is just another variable in the regression model, we can specify the dependence of the hazard on time as any function (e.g. logarithm, quadratic, etc.).

The dependent variable was coded 1 if contact during the current attempt with the sampled person resulted in a completed interview, or 0 if contact did not result in an interview. Independent variables for this analysis included whether the MEPS contact attempt was a similar day and time as the final NHIS interview, whether the contact was the first attempt by the interviewer, the day and time of the contact attempt, the number of days since the last call attempt, the number of true contacts, whether or not the respondent was reluctant to respond, whether the case was ever traced or transferred to a different interviewer, whether or not the respondent shared anything with the interviewer to help get a later interview, region of residence, urbanicity, race and/or ethnicity of the respondent, evidence of non-English speaker in the household, and whether or not the respondent would agree to do a future survey linked to the NHIS.

4. Results

The results presented in Table 3 are from the full model, incorporating a total of 89,930 contact attempts, with 15,447 (17.2 percent) resulting in a completed MEPS interview and 74,483 (82.8 percent) resulting in non-interview. Contact attempts resulting in completed interviews in Round 1 were slightly lower than for Rounds 2 and 3 (15.5 percent versus 18.1 percent). When examining selected measures of the combined Paradata file, Table 4 shows that about 29 percent of Round 1 cases have contact attempts at a similar day and time as the NHIS final interview. In addition, about 14.1 percent of Round 1 cases are reluctant to respond (eventually respond after initially refusing to participate), compared to 4.4 percent in Rounds 2 and 3. Previous analyses have shown that reluctant respondents in the first round of MEPS are significantly more likely to become non-respondents in the second round (Cohen, et. al., 1998).

Table 5 shows the mean number of true contacts by whether the MEPS contact day and time were similar to the final NHIS interview. The mean number of true contacts is slightly lower in Round 1 (8.0 versus 8.6) when the MEPS contact day and time were similar. Also, a higher percentage of Round 1 contact attempts resulted in completed interviews when the day and time were similar (17.7 percent versus 14.6 percent).

After controlling for the demographic and socio-economic factors under study, the model shows the odds of obtaining a completed interview at the next call were 15 percent higher when the MEPS contact day and time was similar to the NHIS final interview (see Table 6). Positive effects were also found when MEPS cases were contacted on weekday evenings, as one would expect. Weekends and weekday evening hours are highly productive for initial contact, and have been found to improve contactability, or the effort involved in making contact with a sampled household (Dahlhamer, et al., 2006). Conversely, contact attempts during weekend mornings reduced the odds of a completed interview. Waiting 3 days after the previous contact attempt, compared to same day next attempts, also increased the odds of obtaining a completed interview by 55 percent.

Not surprisingly, the first contact attempt was associated with lower odds of a completed interview. In addition, MEPS cases that were ever traced, and respondents who were classified as non-Hispanic Other, compared to non-Hispanic white, had decreased odds of a completed interview.

5. Summary

Using the MEPS household component survey, along with the NHIS, we examined whether MEPS contact attempts on a similar day and time as the final NHIS interview were more likely to result in higher first round response and fewer numbers of contacts. We also used available MEPS paradata, to determine some of the predictors that affect the propensity of a case to be interviewed on the next call attempt. Survey research has shown that prior contact with the household is the largest predictor of a successful screener (Groves, et.al., 2005), and our findings indicated that MEPS Round 1 field interviewers may have beneficial outcomes by utilizing contact day and time information available from previous contact with the NHIS. A strong impact was also shown for interviewers who waited 3 days after the previous contact attempt to obtain a completed interview.

Given the findings of this paper, survey operations may consider evaluating efforts to optimize contact attempts, particularly during the first round of the survey. Interviewer training materials should highlight the importance of contacting sample households on a similar day and time as the NHIS final interview as a first round strategy. Our findings show that the percentage of cases contacted on a similar day and time as the final NHIS can be improved (currently only about 29 percent of cases). However any operational

changes must also balance survey response rates as well as data collection costs. Additional analysis using more timely contact history data is warranted.

References

Allison, Paul D. “2010 Survival Analysis Using SAS: A Practical Guide, Second Edition”. Cary, NC: SAS Institute Inc.

Cohen SB, Machlin SR, Branscome J. 1998. “Patterns of Survey Attrition and Reluctant Response in the 1996 MEPS”. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 546-551

Cohen JW, Cohen SB, Bantlin JS. 2009. “The Medical Expenditure Panel Survey—a National Information Resource to Support Healthcare Cost Research and Inform Policy and Practice.” *Med Care* 2009 Jul;47(7 Suppl 1):S44-S50

Dahlhamer, J.M., B.J. Stussman, C.M. Simile, and B. Taylor. 2006. “Modeling Survey Contact in the National Health Interview Survey (NHIS).” Proceedings of the 22nd Statistics Canada International Methodology Symposium.

Groves, R.M., G. Benson, W.D. Mosher, J. Rosenbaum, P. Granda, W. Axinn, J. Lepkowski, and A. Chandra. 2005. “Plan and Operation of Cycle 6 of the National Survey of Family Growth.” Vital and Health Statistics, Series 1, Number 42. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Kreuter, F. and C. Casas-Cordero (2010). “Paradata”. German Council for Social and Economic Data Working Paper Series, Berlin, Germany, No. 136 (April).

National Center for Health Statistics (2004). “Data File Documentation, National Health Interview Survey, 2004 (machine readable data file and documentation)”. National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.

Westat (2007).” MEPS Annual Methodology Report”, No. 74, June 1, 2007. Rockville, MD.

Tables

Table 1. MEPS Data Collection Results for Panel 10, Rounds 1-2

Round	Net Sample RU's	Completes	Average hours/ complete	Response Rate (%)
1	9,086	7,175	11.0	79.0
2	7,381	6,940	7.8	94.0

Source: MEPS Annual Methodology Report, No. 74, June 1, 2007

Table 2. MEPS Summary of Nonresponse for Panel 10, Rounds 1-2

Round	Response Rate (%)	Refusal Rate (%)	Unlocated Rate (%)	Other nonresponse Rate (%)
1	79.0	16.6	3.3	1.1
2	94.0	4.5	0.9	0.6

Source: MEPS Annual Methodology Report, No. 74, June 1, 2007

Table 3. Percent of MEPS Contact Attempts Resulting in Interview, by Round. Panel 10

Round	Completed Interview (%)
Overall	17.2
1	15.5
2-3	18.1

Source: Center for Financing, Access and Cost Trends, AHRQ, Household Component of the Medical Expenditure Panel Survey, 2005, National Health Interview Survey, 2004

Table 4. Selected Paradata Measures by Round, MEPS Panel 10

Paradata	Round 1	Rounds 2-3
MEPS contact day and time same as final NHIS (%)	28.8	27.0
Reluctant respondent (%)	14.0	4.4
Case ever traced (%)	19.0	8.9
Urban (%)	73.8	69.1
Number of true contacts (mean)	8.4	7.3
Number of days since last call attempt (mean)	5.5	4.5

Source: Center for Financing, Access and Cost Trends, AHRQ, Household Component of the Medical Expenditure Panel Survey, 2005, National Health Interview Survey, 2004

Table 5. Contacts by Day and Time, MEPS Panel 10

	MEPS Contact Similar to NHIS	MEPS Contact Different than NHIS
Number of true contacts (mean)		
Overall	7.5	7.8
Round 1	8.0	8.6
Contacts led to interview (%)		
Overall	19.1	16.3
Round 1	17.7	14.6

Source: Center for Financing, Access and Cost Trends, AHRQ, Household Component of the Medical Expenditure Panel Survey, 2005, National Health Interview Survey, 2004

Table 6. Discrete hazard model coefficients and marginal odds-ratios predicting likelihood of a completed interview on the next call attempt: MEPS, 2005

<i>Predictor</i>	<i>Coefficient</i>	<i>p-value</i>	<i>Odds ratio</i>
Intercept	-0.795	<.01	0.45
MEPS contact same day and approximate time as NHIS final contact=1	0.141	<.01	1.15
First attempt=1	-0.713	<.01	0.49
Day and time of contact attempt (ref=Weekend evening)			
Weekend morning/afternoon	-0.154	<.01	0.86
Weekday evening	0.163	<.01	1.18
Weekday morning/afternoon	0.092	<.01	1.10
Anything interviewer should know to help get later interview=1	0.029	NS	1.03
Region of residence (ref=West)			
Northeast	-0.001	NS	1.00
Midwest	-0.012	NS	0.99
South	0.011	NS	1.01
Urban=1	0.163	<.01	1.18
Agree to do future survey linked to NHIS=1	-0.032	NS	0.97
Evidence of non-English speakers=1	-0.008	NS	0.99
Number of true contacts	-0.147	<.01	0.86
Reluctance to respond=1	-0.162	<.01	0.85
Case ever traced=1	-0.674	<.01	0.51
Case ever transferred=1	-0.139	NS	0.87
Race/ethnicity (ref=White, non-Hispanic)			
Hispanic	-0.006	NS	0.99
Black, non-Hispanic	0.022	NS	1.02
Other, non-Hispanic	-0.046	<.05	0.96
Days between previous and current attempt (ref=Same day)			
1 day	0.252	<.01	1.29
2 days	0.315	<.01	1.37
3 days	0.439	<.01	1.55
4 or more days	0.211	<.01	1.23

Notes:

NS= not significant at the .05 level or better.

Source: Center for Financing, Access and Cost Trends, AHRQ, Household Component of the Medical Expenditure Panel Survey, 2005, National Health Interview Survey, 2004