Characteristics of Survey Attrition in the Household Component of the Medical Expenditure Panel Survey (MEPS)[1]

David Kashihara and Trena M. Ezzati-Rice
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

Key Words: Nonresponse bias, weight adjustments, panel survey

Introduction

A common problem for all panel surveys, i.e., surveys with multiple waves or rounds of data collection, is survey attrition. Sometimes respondents drop out of surveys for reasons beyond the control of the data collection agent. In other cases, respondents lose interest in the survey or grow weary from the burden of responding to multiple waves of the survey. Whatever the cause or causes for attrition, the fact remains that it does happen and will continue to happen, despite a survey organization's best efforts to obtain high response rates. Attrition in panel surveys can be a major source of bias if the nonrespondents differ systematically from the respondents with respect to values of the survey variables. This study uses information from the household component of the Medical Expenditure Panel Survey (MEPS-HC) to examine the characteristics of those who initially responded to the survey but failed to complete their full term of eligibility.

The MEPS-HC is a large, national probability sample survey sponsored by the Agency for Healthcare Research and Quality. The survey collects information from respondents to produce annual national- and regional-level estimates of health care use, health status, health conditions, medical expenditures, sources of payment, insurance coverage, and health care access for the U.S. civilian noninstitutionalized population as well as for policy-relevant sub-groups. The MEPS-HC is a two year panel survey with a new panel introduced each year. Five rounds of interviews are conducted with each new panel to yield health care use and expenditure data for two calendar years. The MEPS sample is a sub-sample of respondents to the prior year's National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics. The MEPS sample is drawn from approximately one-half of the PSUs in the NHIS and the oversampling of blacks and Hispanics in the NHIS is carried over to the MEPS (Cohen SB, 2000).

Nonresponse can affect the quality of survey statistics. To counter the potential bias in survey estimates in the MEPS-HC two separate nonresponse adjustments are carried out. The first is an adjustment for dwelling unit (household) nonresponse at Round 1 to account for nonresponse among those households subsampled from NHIS for the MEPS. The second is a person-level nonresponse adjustment to compensate for survey attrition across the various rounds of data collection. This paper deals only with the survey attrition adjustment. MEPS uses a tree algorithm method in the Chi-squared Automatic Interaction Detector (CHAID) to model the response probability at the person-level (Cohen, DiGaetano, and Goksel, 1999). The results are then used to create adjustment cells that are based on the assumption that respondents and drop-outs are similar with respect to the following set of demographic, household, and survey operational characteristics: age, gender, marital status, race/ethnicity, family size, metropolitan statistical area (MSA) type, and indicator for initial refusal to the Round 1 interview (reluctant respondent). This current set of 7 classifying variables is based upon analyses of nonresponse data from the 1996 MEPS and previously described by Cohen and Machlin, 1998 and Cohen *et. al.*,2000. This study examines more recent MEPS data (1999 and 2000) and explores a variety of demographic, socio-economic, household, health, and survey operation variables in an attempt to develop a model that may better reflect the factors related to declining survey participation.

Methods

This study examines those MEPS Panel 4 (1999-2000) participants who initially responded at Round 1 but who failed to complete their full term of eligibility. In this paper, we present two studies: Year 1 attrition and Year 2 attrition. The first study examines the attrition of Round 1 respondents in Rounds 2 and 3 (Year 1 attrition). The second study examines the attrition of Round 3 respondents in Rounds 4 and 5 (Year 2

[1]The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred.

attrition). For both studies, we first assess the characteristics of those who dropped out of the survey for that particular year. A drop-out is defined as an in-scope person who fails to complete their full term of eligibility. This does not include out-of-scope people, i.e., who were admitted to an institution or entered the military, nor does it include those who died during the fielding of the panel. We conducted both bivariate and multi-variate analyses. The final set of predictors of survey attrition from the logistic regression analysis were used to calculate propensity scores.

Study 1 - Year 1 Attrition

Persons in the first study were required to have responded to Round 1 of MEPS Panel 4 in 1999 (n = 14,974). That is, Year 1 attrition was measured as those who responded to Round 1 but who did not respond to either Round 2 or Round 3. The person-level attrition rate for Year 1 was 10.3 percent (Table A).

Table A. Person-Level Nonresponse in MEPS-HC Panel 4 After Round 1: Number and Percent

Sample	Number	Percent
Responded to Round 1	14,974	100.0
Attrited in Year 1	1,542	10.3
Attrited in Year 2*	602	4.5

* *Among Year 1 respondents*

The candidate predictor variables were selected from the MEPS public use files and included the 7 currently used variables and 11 new variables. Table 1 shows the bivariate relationship that each of the 18 Year 1 variables had with respect to Year 1 attrition. The significant variables included not only demographic, household, and survey operation variables, but also a socio-economic measure (employment status of reference person) and a health-related variable (health insurance status). To develop a model for Year 1 attrition, all 18 variables for the bivariate analysis were included in a backwards logistic regression where a greater than 0.05 significance level was required for removal from the final model.

Study 2 - Year 2 Attrition

Year 2 drop-outs were those people who completed Round 3 of Panel 4 (year 1 respondents) but dropped out either in Round 4 or Round 5. The person-level attrition rate for Year 2 was 4.5 percent (Table A). The Year 2 candidate predictor variables included the same set of 18 as in the first study, but also 4 new variables that were not available in the Year 1 data: "total healthcare expenditures in Year 1", "total number of office-based doctor visits in Year 1", "number of medical conditions in Year 1", and "family income as a percent of the poverty line in Year 1". As Table 2 illustrates, different variables were significant with respect to Year 2 and Year 1 attrition, including 3 of the 4 new Year 2 variables. As for Year 1, the model for Year 2 attrition was developed using backwards logistic regression with the same criterion for removal from the final model.

Results

Year 1 Attrition

The bivariate analysis of the 18 candidate predictor variables with respect to Year 1 attrition resulted in 10 variables that exhibited a significant association with Year 1 attrition at the .05 level (Table 1). To develop a model for Year 1 attrition to determine the significant predictors of attrition, all 18 variables were entered into a backwards logistic regression elimination procedure. The model for the predicted probability (p) of Year 1 attrition had the form:

$$\ln[p/(1-p)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{18} X_{18} + \epsilon$$

Table 2 provides a summary of the 9 variables found to be significant predictors of Year 1 attrition. Of the 10 variables that showed a significant bivariate association with Year 1 attrition, only race/ethnicity failed to be included in the final Year 1 model. Among the 7 currently used attrition adjustment variables, 3 were not included in the Year 1 reduced model: marital status, race/ethnicity, and gender. However there were 5 new predictors: health insurance status, education, employment status, race, and region.

A propensity score analysis was carried out to examine the effectiveness of the model. As expected, the attrition rates per quintile, where the first quintile contains those people with the lowest predicted probabilities to drop out and the fifth with those people with the highest probability of dropping out, steadily increased from one quintile to the next (Table B). The attrition rates in the 5^{th} quintile were 3.6 times that of the 1^{st} quintile.

Table B. Distribution of Attrition Rates Based on Predicted Probability for Year 1 Attrition, MEPS-HC, 1999.

Quintile	Attrition Rate (%)
First	4.8
Second	7.6
Third	9.9
Fourth	12.0
Fifth	17.3
Overall (n=1,542)	10.3

Year 2 Attrition

The Year 2 attrition analysis benefitted from the availability of 4 new variables: total healthcare expenditures in Year 1, total number of office-based visits in Year 1, number of medical conditions in Year 1, and family income as a percent of poverty line in Year 1. The Year 2 bivariate analysis of the 22 candidate predictor variables produced 15 variables that showed a significant association with Year 2 attrition at the .05 level (Table 1). To determine the significant predictors of Year 2 attrition, all 22 variables were entered into a backwards logistic regression elimination procedure. The resulting reduced model resulted in the 11 variables shown in Table 3. Of the 7 currently used attrition adjustment variables, 3 were not included in the Year 2 reduced model (gender, race/ethnicity, and MSA) and there were 7 new ones.

As for Year 1, a propensity score analysis provides an illustration of the effectiveness of the model. Again, as expected, the attrition rate increased by quintile and the 5^{th} quintile was about 3 times that of the 1^{st} quintile (Table C).

Table C. Distribution of Attrition Rates Based on Predicted Probability for Year 2 Attrition, MEPS-HC, 2000.

Quintile	Attrition Rate (%)
First	2.7
Second	2.8
Third	3.6
Fourth	5.5
Fifth	7.8
Overall (n=602)	4.5

Discussion

The results from this study indicate that predictors of survey attrition differ between Year 1 and Year 2. Also, some important health and health-related variables were found to be significant and might help in reducing bias in the survey estimates. Since the new models for Year 1 and Year 2 attrition both differ from the current 7 variable model, a Hosmer-Lemeshow goodness-of-fit test (Hosmer and Lemeshow, 2000) was used to compare the new models to the current model. For Year 1, the reduced model had a p-value of 0.094 (chi-square 13.54, 8 df), while the current model had a p-value of 0.004 (chi-square 22.36, 8 df). Thus, for the Year 1 reduced model at the .05 level of significance, we cannot reject the null hypothesis that the new model provides a good fit of the data. However, at the same level of significance, we can reject the null hypothesis for the current model. Therefore, for Year 1 attrition, the new model better explains the data than the current model.

However, the Year 2 attrition goodness-of-fit tests were not as conclusive. The reduced model had a p-value of 0.246 (chi-square 10.28, 8 df), while the current model had a p-value of 0.285 (chi-square 9.72, 8 df). Thus, we cannot reject the null hypothesis that either model provides a good fit to the data. Nevertheless, we are encouraged by the new model because the variables in that model include three salient, health-related variables: self-perceived health status, total healthcare expenditures in Year 1, and number of office-based doctor visits in Year 1.

Future Research

The research presented in this paper is the starting point for several new areas of methodological research to better understand survey attrition in MEPS. For example, one interesting finding from the current research was the observed difference in attrition between the lowest healthcare expenditure group (lower 90 percent in total expenditures) and the highest expenditure group (top 10 percent in total expenditures). Further research is warranted to discern why those with the highest healthcare expenditures tend to drop out of the survey at a higher rate and what can be done to encourage their continued participation in the survey. Other areas of research include the examination of other panels to confirm these results and examination of some observed inconsistencies within year 1 and year 2 in the patterns of attrition by type of health measure. Since the models in this paper usedonly the main effects of the covariates, future research will include investigation of logistic models with interaction terms. In addition, we

will use the new set of covariates to form the nonresponse adjustment cells followed by comparison of estimates and variability of weights based on the current set versus the new set. An evaluation of bias would also be informative. Finally, the weighting research will be extended to an evaluation of the use of propensity scores to adjust for survey attrition. Research on the use of propensity scores to adjust for dwelling unit nonresponse in the MEPS has already been initiated (Wun, *et al.*, 2004).

References

Cohen S.B. (2000), "Sample Design of the 1997 Medical Expenditure Panel Survey Household Component. MEPS Methodology Report No. 11. AHRQ Pub. No. 01-0001. Rockville (MD): Agency for Healthcare Research and Quality.

Cohen S.B., DiGaetano, R., and Goksel, H., "Estimation Procedures in the 1996 Medical Expenditure Panel Survey Household Component." Rockville (MD): Agency for Health Care Policy and Research; 1999. MEPS Methodology Report No. 5. AHCPR Pub. No. 99-0027.

Cohen, S.B., and Machlin, S.R. (1998), "Nonresponse Adjustment Strategy in the Household Component of the 1996 Medical Expenditure Panel Survey," Journal of Economic and Social Measurement, 25, 15-33.

Cohen, S.B., Machlin, S.R. (2000), "Survey Attrition Considerations in the Medical Expenditure Panel Survey," Journal of Economic and Social Measurement, 26, 83-98.

Cohen, S.B., Machlin, S.R., and Branscome, J.M. (2000), "Patterns of Survey Attrition and Reluctant Response in the 1996 Medical Expenditure Panel Survey," Health Services & Outcomes Research Methodology, 1:2, 131-148.

Hosmer, D.W. and Lemeshow, S., Applied Logistic Regression, 2nd Ed., " © 2000 by John Wiley & Sons, Inc., New York, NY, pp. 140-145.

Wun, L., Ezzati-Rice, T.M., Baskin, R., Greenblatt, J., and Zodet, M. (2004), "Using Propensity Scores to Adjust Weights to Compensate for Dwelling Unit Level Nonresponse in the Medical Expenditure Panel Survey," Proceedings of the American Statistical Association, Survey Research Methods Section, [CD-ROM], Alexandria, VA: American Statistical Association.

Table 1. Rate of Attrition for Those Who Responded to Round 1 of MEPS-HC Panel 4 by Response Year.

Measure (no. of dropouts)	**Year 1** (1542)	**Year 2** (602)
Overall	10.30	4.48
Demographic and Socioeconomic Variables		
Age	**	**
< 24 years	10.44	4.06
25 - 34 years	12.21	5.03
35 - 44 years	11.16	5.00
45 - 64 years	9.43	3.40
65 + years	7.82	6.58
Gender		
Male	10.47	4.64
Female	10.14	4.33
Race/Ethnicity	**	
Hispanic	9.13	4.32
Black, non-Hispanic	12.70	4.01
Other	10.19	4.66
Race	**	
American Indian/Aleut/Eskimo	7.69	3.47
Asian or Pacific Islander	18.31	3.36
Black	12.31	4.05
White	9.66	4.61
Marital status of the reference person		
Married	10.12	4.26
Not married or other	10.68	4.96
Education level of the reference person	**	*
No school	3.54	6.42
Grades 1 - 12	9.10	4.46
College	11.88	4.35
Other	13.29	9.49
Employment status of the reference person	**	*
Employed	11.04	4.18
Not employed	8.01	5.28
Other	18.52	6.06
Family income as a % of the poverty line	–	
Poor or negative		4.38
Near poor		4.09
Low income		5.06
Middle income		4.25
High income		4.54

Measure	Year 1	Year 2
(no. of dropouts)	(1542)	(602)
Geographic Variables		
Region	**	**
Northeast	13.08	6.99
Midwest	10.86	3.85
South	8.86	3.91
West	10.03	4.16
MSA	**	**
MSA	11.30	4.78
Non-MSA	6.65	3.46
Household Variable		
Number of people in the RU	**	
1	10.38	4.91
2	10.64	5.12
3	11.95	3.90
4	10.15	4.79
5 +	8.88	3.89
Health-Related Variables		
Insurance coverage	*	
Yes	9.99	4.32
No	11.62	5.18
Perceived health status		**
Excellent	11.01	4.93
Very good	9.94	3.71
Good or other	9.63	4.09
Fair	10.60	5.16
Poor	10.15	8.65
Perceived mental health status		**
Excellent	10.97	4.51
Very good	9.77	3.86
Good or other	9.27	4.61
Fair	11.68	6.20
Poor	10.00	12.50
IADL screener		**
Yes	8.90	10.78
No or other	10.34	4.29
ADL screener		**
Yes	8.89	13.66
No or other	10.32	4.34
Total healthcare expenditures	--	**
$ 0		5.56
$ 1 - 70		5.18
$ 71 - 400		4.23
$ 401 - 1500		3.57
$ 1501 +		4.67

Measure (no. of dropouts)	Year 1 (1542)	Year 2 (602)
Health-Related Variables (continued)		
Number of office-based provider visits	--	**
0		6.14
1		4.63
2 - 4		3.70
5 +		3.12
1999 Total number of health conditions	–	**
0		5.54
1		4.93
2		3.85
3		2.92
4 +		3.39
Survey Operational Variables		
1st Respondent indicator		**
1st Respondent	9.84	3.59
Not 1st respondent	10.59	5.05
Proxy respondent indicator		**
Respondent is an RU member	10.29	4.33
Respondent is a proxy	12.50	3.82
Reluctant respondent	**	**
Yes	21.34	7.80
No	9.34	4.23

* *chi-square test significant at the 0.05 level*
** *chi-square test significant at the 0.01 level*
– *data not available for Year 1*
Sources: Agency for Healthcare Research and Quality, 1999 and 2000 Medical Expenditure Panel Survey - Household Component

Table 2. Year 1 Attrition Variable Reduction Results: Final Model.

Effect	DF	Wald X^2	Pr > X^2
Age	4	18.7	0.001
Race	3	38.5	<.001
Education of ref. psn.	3	12.0	0.007
Empl. status of ref. psn.	2	13.2	0.001
Region	3	23.1	<.001
MSA	1	20.3	<.001
Health insurance status	1	21.0	<.001
No. of people in RU	4	29.3	<.001
Reluctant respondent	1	209.4	<.001

Table 3. Year 2 Attrition Variable Reduction Results: Final Model.

Effect	DF	Wald X^2	Pr > X^2
Age	4	25.5	<.001
Marital status	1	4.2	0.042
Education of ref. psn.	3	9.4	0.025
Region	3	55.6	<.001
Self-perceived health status	4	25.8	<.001
Total healthcare expenditures	4	11.5	0.022
No. of office-based Dr. visits	3	53.5	<.001
First respondent	1	14.3	<.001
Proxy respondent	1	34.0	<.001
No. of people in RU	4	17.7	0.001
Reluctant respondent	1	7.3	0.007