ASSESSMENT OF THE IMPACT OF HEALTH VARIABLES ON NONRESPONSE ADJUSTMENT IN THE MEDICAL EXPENDITURE PANEL SURVEY (MEPS)

Lap-Ming Wun, **Trena Ezzati-Rice**
Department of Health and Human Services, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Abstract

MEPS is a complex sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). Its sample is drawn from respondents to the previous year's National Health Interview Survey (NHIS). The MEPS, like most sample surveys, experiences unit nonresponse. Therefore, the base weight of MEPS respondents is adjusted to compensate for nonresponse. The first level adjustment is at the dwelling unit (DU) level which accounts for nonresponse among those households subsampled from NHIS for the MEPS. The adjustment is done using socio-economic, demographic, and health variables. In this study, we examine the impact of health variables on the MEPS DU level nonresponse weight adjustment. Response propensity scores based on various combinations of covariates are calculated and quintiles of the propensity scores are used to adjust the MEPS base weights for the assessment.

KEY WORDS: logistic regression, response propensity, variation

1. Introduction

The Medical Expenditure Panel Survey (MEPS) is a complex national probability sample survey sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage as well as information on respondents' health status, demographic and socio-demographic characteristics, employment status, access to care, and satisfaction with health care for the U.S. civilian noninstitutionalized population. The MEPS consists of three inter-related surveys with the Household Component (HC) as the core survey. The MEPS-HC, like most sample surveys, experiences unit, or total, nonresponse despite intensive efforts to maximize response rates. Post-survey weighting adjustments are a common tool for reducing the potential bias in survey estimates due to nonresponse. Nonresponse adjustment methods make use of covariates that are available for both survey respondents and nonrespondents. A covariate for a weighting adjustment must have two characteristics to reduce nonresponse bias – it needs to be related to the probability of response, and it needs to be related to the survey outcome (Little and Vartivarian, 2005). The classification tree method, Chi-squared Automatic Interaction Detector (CHAID), is currently employed in MEPS to model the response probability and to form the weighting adjustment cells at the dwelling unit (DU), i.e., household, level (Cohen, DiGaetano, and Goksel, 1999). An alternative method is to calculate response propensity from logistic models based on response related covariates to carry out the nonresponse adjustment as described in Kalton and Flores-Cervantes (2003), and Little (1986). The calculated response propensities can then be used to construct adjustment cells or used directly to adjust the weights. Comparisons of these two alternative methods of response propensity modeling have been studied using earlier panels of MEPS data (Wun *et al*, (2004), and Wun *et a*l, (2007)).

Health and expenditure-related variables are key variables of interest to MEPS data users to inform health policy. To reduce nonresponse bias in the survey estimates, MEPS makes use of a wide variety of available frame variables based on its unique sample linkage to another large ongoing national health survey. In this paper, we examine the effect of nonresponse weighting adjustments that include versus exclude health-related variables, along with other covariates to reduce potential nonresponse bias.

2. MEPS Survey Design and Nonresponse Weighting Adjustments

Each year the sample for the MEPS-HC is drawn from respondents to the previous year's National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention. The MEPS-HC uses an overlapping panel design in which data are collected through a series of five rounds of interviews over a two and one-half year period, and the data from the overlapping panels are used to produce annual estimates. Detailed information on the MEPS sample design has been previously published (Cohen, 1997; Cohen, 2000).

Two separate nonresponse adjustments are made as part of the development of analytic weights in MEPS. The first is an adjustment for dwelling unit (DU) nonresponse at round one to account for nonresponse among those households subsampled from NHIS for the MEPS. The second is a person level nonresponse adjustment to account for survey attrition across the multiple rounds of data collection. This paper reports the findings of research associated with the DU level

nonresponse adjustment.

The base weight in the MEPS is an intermediate weight from the NHIS reflecting the disproportionate sampling of minorities in NHIS with a ratio adjustment to the NHIS population estimates to account for NHIS nonresponse and undercoverage. This ratio adjusted base weight is then adjusted for DU level unit nonresponse in MEPS (i.e., nonresponse among eligible sample DUs at round 1). Specifically, the base weight of MEPS responding DUs is adjusted to compensate for the nonresponding DUs.

3. Methods

The use of classifying or auxiliary variables, i.e., covariates, to form nonresponse adjustment cells is a commonly used method for nonresponse adjustment. It has been shown by Cochran (1968) that it is effective in removing nonresponse bias in observational studies. Rosenbaum and Rubin (1984) have indicated that as the number of covariates increases, the number of classes grows exponentially and suggest using predicted response probabilities or propensity scores from a logistic regression model based on the covariates to form the weighting classes or cells. A propensity score of response in surveys is essentially the conditional probability that a person or household responds given the covariates. Adjustment cells can be formed using the propensity scores, e.g., grouping the sample units (responding and nonresponding units) by quintiles. We use this method in this paper. More elaboration of the propensity score and its application in nonresponse adjustments can be found in Little (1986) and Little and Rubin (2002) among others.

4. Covariates of Response

Since each annual MEPS sample is a subsample of respondents to the previous year's NHIS, a wide range of survey variables for all the MEPS sampled units (respondents and nonrespondents) are available from the MEPS sampling frame, that is, the NHIS. The following set of NHIS variables, identified as relevant to response by Cohen and Machlin (1998), with subsequent updates by Kashihara *et al* (2003) are used as potential covariates for nonresponse adjustment in the MEPS:

1. Age of the reference person
2. Race/ethnicity of the reference person
3. Marital status of the reference person
4. Gender of the reference person
5. Number of persons in the DU
6. Education of the reference person
7. Family income of the reference person
8. Employment status of the reference person
9. Phone number refused in NHIS
10. Major work status – not working due to health reasons
11. DU level health status
12. If anyone in the DU needs help with daily activities
13. Census region
14. Metropolitan Statistical Area (MSA) size
15. MSA/Non MSA residence
16. Urban/Rural residence
17. Type of primary sampling unit (PSU)
18. Predicted poverty status of the household
19. Any Asian in the household
20. Any Black in the household.
21. Interview language
22. US citizenship of the reference person
23. Born in US - reference person
24. Type of home, e.g., house, apartment etc.
25. Time period without phone - interruption in phone service
26. Family medical expenses category
27. Homeowner status of the reference person
28. Number of nights in the hospital last year
29. Healthcare coverage.

There are five health-related variables among these 29 potential nonresponse adjustment covariates: 10) Major work status – not working due to health reason, 11) DU level health status, 12) If anyone in the DU needs help with daily activities, 28) Number of nights in the hospital last year, and 29) Healthcare coverage.

The list of 29 covariates is the potential set of covariates for response. Each year a subset of these are identified via CHAID as significant and used in constructing adjustment cells for nonresponse. For the MEPS Panel 9 (fielded in 2004), round 1 DU level nonresponse adjustment, the following 17 variables from the 29 potential covariates were selected and used in the DU level nonresponse adjustment:

1. Race/ethnicity of the reference person
2. Marital status of the reference person
3. Gender of the reference person
4. Number of persons in the DU
5. Education of the reference person
6. Family income of the reference person
7. Major work status – not working due to health reasons
8. Census region
9. Metropolitan Statistical Area (MSA) size
10. MSA/Non MSA residence
11. Urban/Rural residence
12. Type of primary sampling unit (PSU)
13. Predicted poverty status of the household
14. Any Black in the household.
15. Time period without phone - interruption in phone service
16. Homeowner status of the reference person
17. Number of nights in the hospital last year

Two of these 17 covariates are health-related variables: Major work status– not working due to

health reasons and number of nights in the hospital last year.

5. Nonresponse Adjustments for Evaluation

In this study, we calculate response propensities (propensity scores) using 3 separate sets of predictors (covariates):

(1) The 17 covariates listed above, which were used in the DU nonresponse adjustment for Panel 9, round 1 – this set includes 2 health-related variables.
(2) 15 covariates – which excludes the 2 health-related variables from the 17 variables in (1). That is, no health covariates are included.
(3) 20 covariates – which adds in the 3 additional health-related variables into the set of 17 variables in (1). Thus, this set uses a total of 5 health related variables.

Propensity scores are calculated using the three separate set of covariates, and they are then used to form quintile (5) adjustment cells, which is the number suggested by Cochran (1968) and shown to be optimal in other studies, e.g., Wun *et al* (2004), and Wun *et al* (2007). The base weights of MEPS responding DUs are adjusted within each quintile to compensate for nonresponse.

6. Evaluation Approach and Variables

With the survey linkage between MEPS and NHIS, nonresponding DUs in MEPS were responding units in NHIS. Therefore, both the MEPS responding and nonresponding units have all the survey information from the NHIS. By applying the MEPS base weight to a set of variables from the NHIS, we can obtain a "target" or "true" value, more specifically, an estimate for a variable of interest if no nonresponse had been experienced in the MEPS. The following four NHIS variables were selected to evaluate the impact of health-related covariates on the nonresponse adjustment:

- Dollar-denominated index – which is the dollar amount of expected expenditures derived from qualitative health status (since there are no actual expenditure data in NHIS)
- Doctor visits – whether the household had any member with doctor visits in past 2 weeks
- Limitations – whether the household had any member with limitation in daily activities
- Barriers – whether the household had any member with a barrier to health care due to cost in past 12 months

We apply the three sets of nonresponse adjusted weights (with zero, two, and five health covariates) to each of these four NHIS survey variables to calculate weighted estimates at the DU level. We also calculate an estimate for each of these variables using the MEPS base weight with no nonresponse adjustment (the intermediate NHIS weight). The estimate calculated using the MEPS base weight is labeled as the "target value" and is based on the full MEPS sample – including responding and nonresponding Dus. For each variable, the absolute difference between the "target value" and the estimate calculated using each of the nonresponse adjusted weights is the bias due to that particular adjustment. For the dollar-denominated index, we calculate the total and mean expenditures. For the other three variables, we calculate the estimate of the percent of households that do not have the condition.

7. Evaluation Measures

The aim of the evaluation is to compare estimates that include adjustments for nonresponse (with and without health covariates) to estimates that do not include any nonresponse adjustments (that is, calculated with the base weight). We assess both the bias and the variation of nonresponse adjusted estimates in comparison to the "target values". The measures include the mean square error (MSE), root mean square error (RMSE), and relative RMSE. The MSE used here is not the mean square error in the strict sense in statistics since the bias is defined in terms of a "target value" rather then the expected value as defined earlier.

MSE = square of bias + square of standard error of the estimate
RMSE = square root of MSE
Relative RMSE = RMSE ÷ target value

The resulting statistics for each of the four evaluation variables are given in tables 1-4. Comparisons of the Relative RMSE by subgroups: age, gender, poverty status, and race/ethnicity are given in tables 5.1 to 5.4.

8. Results

The analysis of differences in the estimates based on the three alternative nonresponse adjusted weights and estimates based on a weight without any nonresponse adjustment suggest the potential for nonresponse bias when adjustments are made without the inclusion of health covariates. More specifically, among the four variables evaluated:

- the exclusion of health covariates in the nonresponse adjustment increased the bias and variation, and
- when health covariates were dropped, from the currently used 17 covariates (with two health variables) to a set of 15 covariates with no health variables, the resulting variation (RMSE, relative RMSE) increased.

For three of the four variables (Dollar-Denominated-Index, Doctor Visits, and Limitations), when additional health covariates were included in the nonresponse model, the resulting variation was reduced by

a small amount. For Barriers to Care, the variation increased slightly when all five health variables were included in the adjustment, but the magnitude of increase was less than that from the model which excluded all the health covariates.

These patterns were also consistent across the subgroups studied. It was observed that for three of the four evaluation variables, with the inclusion of health variables, the greatest reduction in relative RMSE was among the white/other group. There was no consistent pattern with respect to poverty status, while for age the inclusion of health variables resulted in the greatest reduction in relative RMSE for persons 45-64 years.

9. Conclusions

In this paper, we examined whether inclusion of health covariates as an auxiliary variable helps to reduce nonresponse bias and variation in weighted survey estimates. Methodologically, the nonresponse bias analysis took advantage of the unique frame variables available for use in the nonresponse adjustment as a result of the survey linkage of the MEPS and the NHIS. Even though the differences were small, in general, excluding health-related covariates in the DU level nonresponse adjustment resulted in estimates with greater bias and estimates with an increase in variation. Similar patterns were observed across subgroups of the population. In this study, we also observed that the magnitude of change (increase or decrease) in variation was smaller by adding more health covariates than from dropping health covariates. Thus, this is a good indication of the success of the currently used 17 covariates that include two health covariates in reducing nonresponse bias in the survey estimates from the MEPS.

References

Cohen, S.B. (1997). "Sample Design of the Medical Expenditure Panel Survey Household Component". *Agency for Health Care Policy and Research, MEPS Methodology Report, No. 2*, AHCPR Pub. No. 97-0027, Rockville, MD, 1997

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

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

Cochran, W.G. (1968). "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies," *Biometrics*, 24, 295-313.

Kalton, G. and Flores-Cervantes, I. (2003) "Weighting Methods," *Journal of Official Statistics*, Vol. 19, No. 2, 81-97.

Kashihara, D., Ezzati-Rice, T.M., Wun, L-M., and Baskin (2003). "An Evaluation of Nonresponse Adjustment Cells for the Household Component of the Medical Expenditure Panel Survey (MEPS)". *2003 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association: 3758-3765.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis With Missing Data* (2nd Ed.). New York: Wiley.

Little, R.J.A. (1986). "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review*; 54: 139-157.

Little, R.J., and Vartivarian, S., (2005). "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology*, Vol. 31, No. 2, 161-168.

Rosenbaum, P.R., and Rubin, D.B., (1984). "Reducing Bias in Observational Studies Using Subclassification on Propensity Score," *Journal of the American Statistical Association*, 79, 516-524

Wun, L.-M., Ezzati-Rice, T.M., Baskin, R., Greenblatt, J., Zodet, M., Potter, F., Diaz-Tena, N., and Touzani M., (2004). "Using Propensity Scores to Adjust Weights to Compensate for Dwelling Unit Level Nonresponse in the Medical Expenditure Panel Survey," *2004 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association: 4625-4631.

Wun, L.-M., Ezzati-Rice, T.M., Diaz-Tena, N., and Greenblatt, J. (2007). "On Modeling Response Propensity for Dwelling Unit (DU) Level Nonresponse Adjustment in the Medical Expenditure Panel Survey (MEPS)," *Statistics in Medicine*, 2007: 26: 1875-1884.

Table 1. Summary Statistics: Dollar-Denominated-Index (proxy of expenditures): nonresponse (NR) adjusted estimates compared to target value

Weight	Total	Mean	SE of mean	Bias of mean	RMSE	Relative RMSE
MEPS base weight DUPSWT (target value)	767,984,674,958	6,618	115	NA	NA	NA
NR adj. - 20 covariates (all 5 health covariates)	775,107,636,074	6,680	75	62	97	1.46%
NR adj. – 17 covariates (2 health covariates)	775,370,469,420	6,682	75	64	98	1.48%
NR adj. – 15 covariates (no health covariate)	778,845,236,930	6,712	76	94	120	1.82%

Table 2. Summary Statistics: Doctor Visits: nonresponse (NR) adjusted estimates compared to target value

Weight	Percent of DUs with no doctor visit	SE	Bias	RMSE	Relative RMSE
MEPS base weight DUPSWT (target value)	67.23	1.40	NA	NA	NA
NR adj. - 20 covariates (all 5 health covariates)	65.83	0.68	1.40	1.57	2.32%
NR adj. – 17 covariates (2 health covariates)	65.79	0.68	1.44	1.59	2.37%
NR adj. – 15 covariates (no health covariate)	65.67	0.68	1.56	1.70	2.53%

Table 3. Summary Statistics: Limitations: nonresponse (NR) adjusted estimates compared to target value

Weight	Percent of DUs with no member with a limitation	SE	Bias	RMSE	Relative RMSE
MEPS base weight DUPSWT (target value)	74.64	1.50	NA	NA	NA
NR adj. - 20 covariates (all 5 health covariates)	73.06	0.79	1.58	1.77	2.37%
NR adj. – 17 covariates (2 health covariates)	73.05	0.80	1.59	1.78	2.38%
NR adj. – 15 covariates (no health covariate)	72.74	0.80	1.90	2.06	2.76%

Table 4. Summary Statistics: Barriers to Care: nonresponse (NR) adjusted estimates compared to target value

Weight	Percent of DUs with no member with barriers to care	SE	Bias	RMSE	Relative RMSE
MEPS base weight DUPSWT (target value)	85.28	1.04	NA	NA	NA
NR adj. - 20 covariates (all 5 health covariates)	84.32	0.58	0.96	1.12	1.32%
NR adj. – 17 covariates (2 health covariates)	84.37	0.58	0.91	1.08	1.27%
NR adj. – 15 covariates (no health covariate)	84.28	0.58	1.00	1.16	1.36%

Table 5.1. Weighted Dollar-Dominated Index: relative RMSE by subgroups

Sub-group	20*		17*		15*
	(all 5 health)		(2 health)		(no health)
age < 24	5.08%	<	5.09%	>	5.02%
age 24-44	2.39%	>	2.35%	<	2.60%
age 45-64	1.98%	>	1.92%	<	2.29%
age ≥ 65	1.96%	>	1.93%	>	1.88%
male	1.79%	<	1.84%	<	2.10%
female	1.69%	>	1.67%	<	2.01%
with poverty	2.17%	=	2.17%	>	2.06%
No poverty	1.37%	<	1.41%	<	1.64%
Hispanic	3.60%	<	3.64%	<	3.92%
Black	3.28%	<	3.36%	<	3.53%
Other Race	1.39%	<	1.40%	<	1.65%

*

20: model with 20 covariates (17 used in CHAID + 3 additional health covariates)
17: model using the 17 covariates currently used in CHAID
15: model excluding the 2 health covariates from the 17 currently used in CHAID

Table 5.2. Doctor Visits: relative RMSE by subgroups

Sub-group	20*		17*		15*
	(all 5 health)		(2 health)		(no health)
age < 24	3.27%	=	3.27%	=	3.27%
age 24-44	2.64%	<	2.67%	<	2.75%
age 45-64	2.74%	<	2.79%	<	3.01%
age ≥ 65	3.19%	<	3.22%	<	3.36%
male	2.30%	<	2.37%	<	2.48%
female	2.66%	<	2.67%	<	2.85%
with poverty	2.76%	<	2.83%	<	3.00%
No poverty	2.34%	<	2.39%	<	2.53%
Hispanic	2.51%	<	2.56%	<	2.63%
Black	3.37%	=	3.37%	<	3.44%
Other Race	2.49%	<	2.55%	<	2.72%

*
20: model with 20 covariates (17 used in CHAID + 3 additional health covariates)
17: model using the 17 covariates currently used in CHAID
15: model excluding the 2 health covariates from the 17 currently used in CHAID

Table 5.3. Limitations: relative RMSE by subgroups

Sub-group	20*		17*		15*
	(all 5 health)		(2 health)		(no health)
age < 24	1.82%	>	1.80%	<	1.81%
age 24-44	2.15%	>	2.11%	<	2.30%
age 45-64	3.09%	>	3.04%	<	3.63%
age ≥ 65	3.62%	<	3.70%	<	4.04%
male	2.27%	<	2.33%	<	2.64%
female	2.69%	>	2.67%	<	3.06%
with poverty	3.73%	<	3.75%	<	4.46%
No poverty	1.86%	<	1.90%	<	2.14%
Hispanic	2.34%	>	2.32%	<	2.55%
Black	3.45%	<	3.57%	<	3.81%
Other Race	2.41%	>	2.40%	<	2.80%

*
20: model with 20 covariates (17 used in CHAID + 3 additional health covariates)
17: model using the 17 covariates currently used in CHAID
15: model excluding the 2 health covariates from the 17 currently used in CHAID

Table 5.4. Barriers to Care: relative RMSE by subgroups

Sub-group	20*		17*		15*
	(all 5 health)		(2 health)		(no health)
age < 24	3.21%	<	3.27%	>	3.25%
age 24-44	1.65%	>	1.63%	<	1.69%
age 45-64	1.85%	>	1.78%	<	1.91%
age ≥ 65	0.88%	>	0.85%	<	0.89%
male	1.11%	>	1.06%	<	1.15%
female	1.73%	>	1.69%	<	1.76%
with poverty	2.13%	>	2.09%	<	2.15%
No poverty	1.17%	>	1.12%	<	1.18%
Hispanic	1.95%	>	1.89%	<	1.90%
Black	2.17%	<	2.22%	<	2.25%
Other Race	1.22%	>	1.17%	<	1.26%

*

20: model with 20 covariates (17 used in CHAID + 3 additional health covariates)
17: model using the 17 covariates currently used in CHAID
15: model excluding the 2 health covariates from the 17 currently used in CHAID