The Impact of Medical Expenditure Predictors in the Medical Expenditure Panel Survey (MEPS) Nonresponse Adjustment

Lap-Ming Wun, Trena M. Ezzati-Rice, Steven B. Cohen, William Yu
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850

KEY WORDS: weighting; CHAID; high medical expenditure predictor and indicator; Dollar-dominated index

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, ongoing since 1996, is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage 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. Survey nonresponse is usually compensated for by some form of weighting adjustment to reduce the potential bias in survey estimates. 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. The current method implemented by Westat (Cohen, DiGaetano, and Goksel, 1999) to compensate for nonresponse in the MEPS at the dwelling unit (DU) level uses the Chi-squared Automatic Interaction Detector (CHAID) "tree algorithm" response propensity approach (Breiman, Friedman, Olshen, and Stone, 1993) to form nonresponse adjustment cells.

In this paper, the current set of covariates used in MEPS to adjust for DU level nonresponse is supplemented with an additional covariate, a predictor of high medical expenditures. We discuss the model used to construct the high medical expenditure predictor variable and how it is used in the nonresponse adjustment. We then evaluate the impact of the inclusion of this new variable. This study is done using MEPS panel 7 data (new panel of 2002 only).

Background: MEPS Survey Design and Estimation Strategy

The annual 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 (NCHS). 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. Detailed information on the MEPS sample design has been previously published (Cohen, 1997; Cohen, 2000).

Two separate nonresponse adjustments are performed as part of the process for development of analytic weights in MEPS. The first is an adjustment for DU nonresponse at round 1 to account for nonresponse among those households subsampled from NHIS for the MEPS. The 1996 to 2002 MEPS DU response rates ranged from 80-83 percent (among the NHIS households fielded for MEPS). The second nonresponse adjustment is at the person level to account for survey attrition across the various rounds of data collection. This paper deals only with the DU nonresponse adjustment.

The base weight in the MEPS is the reciprocal of an intermediate weight from the NHIS reflecting the disproportionate sampling of minorities in NHIS with a ratio adjustment to NHIS population estimates to account for NHIS nonresponse and undercoverage. This ratio adjusted base weight is then adjusted for nonresponse among MEPS eligible sample DUs at round 1. More specifically, the base weights of MEPS responding DUs are adjusted to compensate for the nonresponding DUs.

Nonresponse Weighting Adjustment in MEPS

In the method currently used for MEPS, Westat uses a tree diagram generated by the computer package CHAID to form nonresponse adjustment cells based on response propensity using a set of classifying variables. Cells are collapsed, if necessary to ensure that the number of respondents in a cell is no less than 20 (Göksel, Alvarez-Rojas, and Hao, (2001)). Adjustment factors are not permitted to exceed two in value in order to limit the impact of such factors on the variability of sample estimates. It should be noted that because of the unique sample linkage of MEPS and the NHIS, a sizeable number of variables are available from the NHIS for responding and non-responding eligible MEPS DUs. The following

is the list of NHIS variables used as potential predictors of response propensity to construct subclasses for the DU nonresponse adjustment in MEPS-HC through 2003. These classifying variables were determined based on analysis of 1996 MEPS-HC data (Cohen and Machlin, 1998) and a follow up analysis of 2000 MEPS-HC data (Kashihara, et al (2003)).

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 – working or reason for not working
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.

The following additional covariates were added in 2004 for the 2002 MEPS:

1. Interview language
2. US citizenship of the reference person
3. Born in US - reference person
4. Type of home, e.g., house, apartment etc.
5. Time period without phone
6. Family medical expenses category – 0, 1 to 500, 500 to 1999, 2000 or more.
7. Homeowner status of the reference person
8. Number of nights in the hospital last year
9. Healthcare coverage (Health insurance).

In this study, we evaluate the impact of a constructed variable, high medical expenditure predictor, on the adjustment of DU level nonresponse by splitting each of the cells used in round 1 of the 2002 MEPS (panel 7) into two, according to the two categories of this new predictor variable.

High Medical Expenditures Predictor and Indicator

High medical expenditures are important factors in many studies using MEPS data. Therefore, expenditure variables are important for data collection as well as nonresponse adjustment. However, NHIS does not collect expenditure data; hence, a predictor using a logistic model based on relevant covariates was developed to predict a person's probability of incurring high medical expenditures in the year subsequent to the NHIS interview. Each person is then given an index indicating whether he/she is likely to be a high expenditure person according to the predicted probability.

The prediction model was originally developed by Moeller and Mathiowetz (1994) using 1987 National Medical Expenditure Survey (NMES, the predecessor of MEPS) data. The model was up-dated using 1997 NHIS and 1998 MEPS data by Wun (2002).

The dependent variable for the model is defined as follows:

> High expenditure is defined as the dollar amount of medical expenditures at the person level in the top 10 percent of the medical expenditure distribution. A person with high expenditures is given a 1 to this dependent variable and a 0 otherwise.

Covariates (independent variables) in the model include:

1. Gender
2. Health status
3. Census division
4. Metropolitan Statistical Areas (MSA)
5. Marital status
6. Poverty status
7. Whether the person lives alone
8. Age
9. Health limitation – kept from work
10. Health limitation in amount or type of work
11. Number of ambulatory visits
12. Number of prescriptions.

With these covariates and the dependent variable as defined above, each person's probability of incurring high medical expenditures is calculated through a logistic regression model. Using that probability, each person is given four high medical expenditure indices. If a person's predicted probability is in the top XX percent (the four values selected for XX are: 5, 10, 15, and 20) of all sampled persons' probabilities, then the person is designated a high medical expenditure person and is given a 1 on the indicator. Denoting the indicator by hiexpXX, hiexp05=1 if the person's probability of incurring high medical expenditure is in the top 5 percent of the distribution of calculated probabilities, otherwise, hiexp05=0. Therefore, each person has four indicators: hiexp05, hiexp10, hiexp15, and hiexp20; each with a value of 1 or 0.

Dwelling Unit (DU) Level High Medical Expenditure Indicator for Nonresponse Adjustment

In order to carry out the DU level nonresponse adjustment evaluation, it was necessary to translate the person level high medical expenditure indicator to the DU level. Thus, if one or more members of a DU have hiexpXX = 1, then the DU is designated a high expenditure DU and given an index value of 1 for the variable DUhiexpXX. Therefore, each DU also has 4 high medical expenditure indicators: DUhiexp05, DUhiexp10, DUhiexp15, and DUhiexp20. Each has a value of 1 or 0.

DU Level Nonresponse Adjustment with High Expenditure Indicator

With the DUhiexpXXs for each DU, the analysis of their impact on nonresponse adjustment is carried out in three ways which we designate as Group A, Group B, and Group C defined as follows:

Group A:

> Uses only the high expenditure indicator, DUhiexpXX, as the adjustment factor. The entire sample falls into only one of two adjustment cells: DUhiexpXX=1 or DUhiexpXX=0. Adjustment is done within each cell.

Group B (rule of 20):

> Split each of the adjustment cells used in round 1 of the 2002 MEPS into two cells according to DUhiexpXX=1 or 0. But the new cells may be collapsed to assure that each cell has at least 20 respondent units (rule of 20).

Group C (rule of 10 or 100%):

> Split each of the adjustment cells used in round 1 of the 2002 MEPS into two cells according to DUhiexpXX=1 or 0, and collapse the resulting cells so that each cell has at least 10 respondent units or all units in the cell are respondents (rule of 10 or 100%).

Within each adjustment cell, the weights of respondent DUs are multiplied by the factor A:

$$A = \frac{\sum_{i \in c} W(i)}{\sum_{i \in c} R(i)W(i)}$$

where A is the ratio of the sum of weights of all units in the cell to the sum of weights of only the respondents (R) in the cell.

DU Level Nonresponse Adjusted Weights

With 4 high expenditure indicators in each of the three adjustment groups, there are 12 sets of nonresponse adjusted weights that are evaluated. All adjusted weights equal to the sum of the MEPS base weights (weights of the respondent as well as nonrespondent units before adjustment). All 12 sets of weights along with the nonresponse adjusted weight currently used in MEPS have the same total, thus the same mean of 16,395. First, we examine the impact on the variation in the weights adding the high medical expenditure indicator into the adjustment for nonresponse. The following table shows the standard deviations of the DU level nonresponse adjusted weights currently in MEPS and the 12 new sets of adjusted weights:

Table 1: Mean Dwelling Unit Weights

Current MEPS	10,051	10,051	10,051

	Group A	Group B	Group C
XX% = 5%	9,382	10,064	10,068
XX% = 10%	9,397	10,060	10,070
XX% = 15%	9,376	10,068	10,066
XX% = 20%	9,371	10,058	10,060

We next used each of the 12 sets of new weights to estimate the mean medical expenditures of all individuals represented by the sample of MEPS panel 7 as a further evaluation of the impact of adding the high medical expenditure indicator in the nonresponse adjustment. However, since NHIS does not collect expenditure data, and MEPS does not have data for nonrespondents, we adopt a predicted measure for this assessment.

Dollar-Dominated Index

A proxy for the dollar amount of medical expenditures for each sampled individual, designated dollar-dominated index of health status, was developed by Selden (2006). It is a measure of expected expenditures derived from qualitative health status. This dollar-dominated index is available for each MEPS sample person - respondent as well as nonrespondent. With this measure as each individual's total medical expenditures, we can calculate an estimate of the mean expenditures of all sampled DUs using their MEPS base weights. This estimate is calculated for the full MEPS sample and it is considered the target value for comparison of estimates using other sets of new weights for this analysis. The difference between each of the estimated means using the new sets of weights and the target estimate is considered as the bias due to the particular nonresponse adjustment. This bias along with the standard error (SE) of the estimate give the root mean square error (RMSE) of each of the estimated means. The results are given in the following table:

Table 2: Summary Statistics for Mean Dollar-Dominated Indices

Type of Weight	Mean	SE	RMSE
MEPS base weight (no NR adjustment)	6699.97	76.42	
Current MEPS NR adjusted weight	6720.35	69.95	72.86
New adjusted weights - Group A			
5%	6777.28	69.50	103.96
10%	6763.46	69.15	93.88
15%	6752.61	68.93	86.73
20%	6746.50	68.84	83.09
New adjusted weights - Group B			
5%	6719.26	69.78	72.40
10%	6725.03	70.03	74.38
15%	6712.13	69.88	70.93
20%	6710.00	69.74	70.46
New adjusted weights - Group C			
5%	6715.28	69.73	71.39
10%	6722.75	70.10	73.71
15%	6702.61	69.56	69.61
20%	6704.44	69.53	69.67

Summary

The results of this study showed only small differences in the standard deviations of the nonresponse adjusted weights by adding a high medical expenditure indicator as an additional factor in adjusting weights to compensate for DU level nonresponse. A small, but not significant, decrease was observed for Group A.

There were some improvements in terms of reduction in RMSE in the estimated mean expenditures using the new weights from Groups B and C, with cut off points of 15% and 20% as compared to the RMSE of the mean estimated using the current MEPS nonresponse adjusted weights. However, the differences were small.

In summary, the small differences observed in this study could be due to the fact that many of the covariates used in calculating the probability of high expenditures are the same variables as used in the current nonresponse adjustment; therefore, most of the effect of correcting for potential nonresponse bias has already been accounted for. With these results, the current nonresponse adjustment in MEPS appears to be successful in correcting for potential nonresponse bias.

An additional issue is that the dollar-dominated index used as proxy for the dollar amount of medical expenditures may not properly represent expenditures of nonrespondents. This will be investigated in a future study.

References

Brieman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1993). Classification and Regression Trees. New York: Chapman and Hall.

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

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.

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, Rockville, MD., AHRQ Pub. No.01-0001.

Cochran, W.G. (1968). "Removing Bias in Observational Studies," Biometrics, 24, 295-313.

Kashihara, D., Ezzati-Rice, T.M., Wun, L.-M., and Baskin, R.M. (2003). "An Evaluation of Nonresponse Adjustment Cells for the Household Component of the Medical Expenditure Panel Survey (MEPS)". 2003 ASA Proceedings of the section on Survey Methodology, pp.3758-3765, (CD-ROM).

Göksel, H.A., Alvarez-Rojas, L., Hao, H., (2001). "MEPS Panel V Round 1 – DU level weights" Unpublished internal memo of Westat, WGTS # 469.01, and P5R1 #1.01.

Moeller, J., Mathiowetz, N. "A Problems of Screening for Poverty Status," Journal of Official Statistics, 1994, Vol. 10, No. 1, pp. 327-337.

Selden, T.M. (2006). "An Expenditure-Weighted Index for Self-Reported Health," Unpublished report, Center for Financing, Access, and Cost Trends (CFACT), Agency for Healthcare Research and Quality (AHRQ).

Wun, L.-M. (2002), ``Refining the Prediction Models of Low Income Status and Future Medical Expenditures in the Medical Expenditure Panel Survey", ASA Proceedings of the Joint Statistical Meetings, 3813-3818.