Using Propensity Scores to Adjust Weights to Compensate for Dwelling Unit Level Nonresponse in the Medical Expenditure Panel Survey
Lap-Ming Wun, Trena M. Ezzati-Rice, Robert Baskin, Janet Greenblatt, Marc Zodet, Frank Potter, Nuria Diaz-Tena, and Mourad Touzani
October 2004

ABSTRACT

The Medical Expenditure Panel Survey (MEPS) is sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS, a complex national probability sample survey, is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. It comprises three component surveys with the Household Component (HC) as the core survey. The MEPS-HC, like most sample surveys, experiences unit nonresponse despite efforts to maximize response rates. Survey nonresponse is usually compensated for by some form of weighting adjustment to reduce the bias in survey estimates. Currently, a weighting class nonresponse adjustment using socio-economic and demographic variables to create the weighting classes is used in the MEPS to adjust for potential nonresponse bias at the dwelling unit level. An alternative method for forming nonresponse adjustment cells is to use response propensities. This paper summarizes research undertaken to investigate various potential use of response propensities to adjust weights to compensate for nonresponse in the MEPS. Survey estimates for selected survey components, CVs, distribution of weights, and nonresponse adjustment methods at the dwelling unit level are compared and methodological issues discussed.

Lap-Ming Wun
Statistician, Center for Financing, Access, and Cost Trends
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
E-mail: LWun@ahrq.gov

Trena M. Ezzati-Rice, Robert Baskin, Janet Greenblatt, Marc Zodet
Center for Financing, Access, and Cost Trends
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

Frank Potter, Nuria Diaz-Tena, and Mourad Touzani
Mathematica Policy Research, Inc.

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 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. Nonresponse adjustment methods make use of covariates that are available for both respondents and nonrespondents. Currently, the tree algorithm method using the Chi-squared Automatic Interaction Detector (CHAID) is employed in MEPS to model the response probability at the dwelling unit (DU), i.e., household, level and to create the nonresponse adjustment cells (Cohen, DiGaetano, and Goksel, 1999). The other adjustment carried out in the development of weights in MEPS is poststratification of the nonresponse adjusted weights to known population counts, i.e., estimates of the population from the Current Population Survey (CPS).

In this paper, we use logistic regression modeling to predict the probability of response using the same set of covariates currently used in CHAID. The nonresponse adjusted weights are then poststratified at the family level to match the CPS control totals. Family level survey estimates for selected survey components, standard errors, relative standard errors, and distributions of weights based on the two alternative propensity nonresponse approaches (CHAID and logistic regression) are compared and methodological issues discussed.

Background: MEPS Survey Design and Estimation Strategy

The sample for the MEPS-HC is drawn from respondents to the National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics. 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 2000 MEPS DU response rates ranged from 80-83 percent (among the NHIS households fielded for MEPS). The second is a person level nonresponse adjustment 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 of 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

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. Another adjustment method is to use the inverse of the respondent's predicted propensity score as an adjustment factor (see Kalton and Flores-Cerantes (2003)). In this paper we call this latter method the "direct use" of propensity scores. A propensity score of response in surveys is essentially the conditional probability that a person or household responds given the covariates. 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. A previous comparison of the use of covariates versus the use of response propensities to form classes for nonresponse adjustment for a complex sample survey, the third National Health and Nutrition Examination Survey (NHANES III), was reported by Ezzati-Rice and Khare (1994).

The current method implemented by Westat to compensate for nonresponse in the MEPS at the DU level uses CHAID's "tree algorithm" response propensity approach (see Breiman, Friedman, Olshen, and Stone, 1993) to form nonresponse adjustment cells. In this research study, we investigate two alternative ways of using response propensities from logistic regression modeling to adjust weights to compensate for nonresponse.

Methods

In this study, we assess the differences among the various methods of DU nonresponse adjustment at round 1 of the 2000 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 17 classifying variables. Cells are collapsed, if necessary to ensure that the number of respondents in a cell are no less than 20 (Göksel, Alvarez-Rojas, and Hao, 2001). 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 17 variables used by Westat as potential predictors of response propensity to construct subclasses for the DU nonresponse adjustment in the 2000 MEPS-HC. These classifying variables were determined based on analysis of 1996 MEPS-HC data (Cohen and Machlin, 1998).

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)

An alternative to the current CHAID propensity nonresponse adjustment method is to develop a logistic regression model to predict response status using the same set of 17 covariates identified above. A propensity score of response in surveys is essentially the conditional probability of response given the covariates. For this study, it was calculated through the following steps:

1. Run a logistic regression with response/nonresponse indicator as the dependent variable using the 17 significant covariates described above.

2. Convert the estimated logit value obtained from the logistic model established in step 1 into the predicted probability of response, i.e., the propensity score, through the following equation:

PROB=EXP(LOGIT)/(1+EXP(LOGIT)).

With a propensity score calculated for each sample unit, the propensity score from the logistic regression is used in two different ways:

1. Direct:

The estimated propensity score of each respondent is used directly as the adjustment factor, i.e., each individual respondent's base weight is multiplied by the inverse of their propensity score.

2. Grouping scores to form adjustment cells:

Using the propensity scores, the sample is grouped into classification cells. In this study, we used groupings of 5 and 100 and compare the results with those from the current CHAID method. The selection of 5 groups was based on the optimality established by Cochran (1968) and extended to propensity scores in observational studies by Rosenbaum and Rubin (1984). These studies showed that 5 classes were often sufficient to remove 90% of the bias due to the covariates. The inclusion of 100 groups was designed to assess the effect of a much finer classification of the propensity scores while keeping the number of respondents in a cell at no fewer than 20 to match the criterion used by Westat in the current method.

Adjustment, Poststratification, and Evaluation Approach

Three methods of constructing weighting classes were evaluated: 1) the current CHAID approach, 2) 5 groups based on predicted response propensities from logistic regression, and 3) 100 groups based on predicted response propensities from logistic regression. After sample DUs were classified into cells by one of the three methods, the base weights of responding DUs were

8

inflated by an adjustment factor within each cell. The adjustment factor was the ratio of the sum of the base weights of all units in the cell to the sum of the base weights of only the respondents in the cell. The direct method of using propensity scores was carried out by multiplying the respondent's base weight by the inverse of its propensity score.

The DU nonresponse adjusted weights based on each of the nonresponse adjustment methods (cell weighting and direct use) were then poststratified at the family level to totals obtained from the March 2000 CPS. The poststratification was done within classes formed by family type, race/ethnicity, region, MSA status, age, and family size. The poststratification may add an element that may confound the comparisons slightly, but was done to permit the development of MEPS estimates for comparison purposes. It is not thought that the findings will be substantially affected by this approach.

The nonresponse adjusted and poststratified weights were applied to selected health and expenditure variables to calculate family level nonresponse adjusted estimates. The resulting distributions of the weights and weighted summary statistics of selected analytical variables from the four nonresponse adjustment approaches (CHAID and the three logistic regression methods) were compared. The weighted estimates shown in this paper are for research purposes only and do not reflect the full set of adjustments undertaken in the survey to produce national estimates.

The analytical variables selected for this research were: insurance coverage, health status, and total health care expenditures (payments).

Insurance coverage was defined into 3 categories:

1=all members in the family were insured (for approximately first half of the year),

9

2 = some of the members were insured (for approximately first half of the year),

3 = none of the members were insured (for approximately first half of the year).

Health status was also defined into 3 categories:

1=all members in the family were reported as in fair or poor health at the round 1 interview,

2 = some members were reported as in fair or poor health at the round 1 interview,

3=all members were reported as in good to excellent health at the round 1 interview.

The total health care expenditures variable reflects total payments for health care services during the year for all persons in the family.

Results

In evaluating the effectiveness of approaches to nonresponse adjustment, one should look at both measures related to bias reduction as well as the increase in variance associated with corresponding increased variation in the weights. Due to time constraints, for this paper we focus only on the latter.

There were 5357 DUs in the first round of panel 5 (the 2000 MEPS panel), of which 4334 responded to the survey. There were 4565 families in these 4334 DUs. The comparison of adjusted weights based on the four adjustment methods with the 17 covariates as currently used by Westat was done based on these 4565 families.

Table 1 shows the standard deviation, mean, minimum, and maximum of the weights after adjustment for DU nonresponse using each of the four methods of adjustment and the poststratification at the family level to the CPS control totals. The standard deviations of the weights have the following order: Direct < 5 groups < CHAID < 100 groups.

Table 2 shows the weighted proportion in each category of insurance coverage using each of the four different weights. For all three categories of insurance coverage, direct use of the propensity score has the smallest standard error, followed by that of the 5 group.

Table 3 shows the weighted proportion in each category of health status using each of the four weights. The results were the same as for health insurance coverage. Specifically, for all three categories of health status, the direct use of the propensity score had the smallest standard error, followed by the 5 groups.

Tables 4 (A) and (B) show the mean and median and their standard error (SE) and relative standard error of the total family health care expenditures under each of the four methods. The size of the relative standard errors of the mean have the following order: Direct < 5 group < CHAID < 100 group. The relative standard errors of the median have the same order.

Discussion

The results reported above are based on an unweighted logistic regression model using the 17 covariates. A logistic model with weights was also investigated. The "weighted" approach follows what Little and Vartivarian (2003) suggest, namely that the weights be included as a covariate in the logistic model rather than running a traditional weighted regression. The weights used in our logistic model were the MEPS base weights.

The regression coefficients and their significance levels (the p-values) for all 17 covariates from either model have very similar values. The coefficient of the weight itself is marginally significant (with a p-value of 0.0929). The p-value of the Hosmer-Lemeshow goodness-of -fit of the model with weights is 0.2356 as compared to 0.8298 for the model without weights. This indicates that the model with weights included does not fit the observations as well as the one without weights. Table 5 provides the resulting distribution of weights with adjustment for nonresponse from the model that includes the base weight. Comparing Table 5 with Table 1, we see that the distribution from the direct method has a slightly wider range in Table 5 than that in Table 1. Also the results from the 5 groups and 100 groups have greater standard deviations in Table 5 than those in Table 1. Therefore, including base weights in the logistic model is not very beneficial at this stage of our research.

Summary

- In general, the logistic regression approaches produce results similar to CHAID;

- Results observed in this study are consistent with other nonresponse adjustment studies;

- Direct use of the propensity scores performed best in terms of limiting variability, followed by the 5-group stratification;

- These two logistic regression-based propensity score methods performed better in terms of the relative variance of the estimates than the CHAID-based method, but the differences were marginal. Additional refinement of stratification in the logistic regression approach, i.e., the 100 groups studied here, increased variation slightly. The relative benefits of the four approaches related to bias reduction remain to be investigated.

- The inclusion of survey base weights in the logistic regression model did not provide any substantial advantage.

Future Research

To continue to assess the difference between the current CHAID method of nonresponse adjustment and that based on logistic regression-based propensity scores, we will carry forward the DU level nonresponse adjusted, family level poststratified weights through subsequent steps of the weights development in MEPS. Those steps include poststratification at the person level, adjustment to account for survey attrition, and final poststratification to match the December 2000 CPS estimates. Since the logistic model in this study used only the main effects of the covariates, future research will include investigation of logistic models with interaction terms. In addition to investigation of variability of weights resulting from alternative nonresponse adjustment methods, an evaluation of bias would also be informative.

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.

Ezzati-Rice, T.M., and Khare, M. (1994). "Modeling of Response Propensity in the Third National Health and Nutrition Examination Survey," 1994 Proceedings of Survey Research Methods Section of the American Statistical Association, 955-959.

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, P5R1 # 1.01.

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)". Paper presented at the 2003 Joint Statistical Meetings, San Francisco, CA., August, 2003.

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

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

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

Little, R.J. and Vartivarian, S., (2003) "On Weighting the Rates in Non-response Weights," Statistics in Medicine, 22, 1589-1599.

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.

Smith, P.J., Rao, J.N.K., Battaglia, M.P., Ezzati-Rice, T.M., Daniels, D., and Khare, M., (2001), "Compensating for Provider Nonresponse Using Response Propensities to Form Adjustment Cells: The National Immunization Survey," Vital and Health Statistics, Series 2, No. 133. DHHS Publication No. (PHS) 2001-1333, Series 2, No. 133.

Table. 1. Distribution of DU level Nonresponse Adjusted and Family level Post-stratified Weights
(Direct and 5 and 100 groups based on unweighted logistic model)

NR Adj. Method*	Standard deviation	Mean	Max	Min
Current Method	13,249	25,534	121,915	2,549
Direct	12,614	25,534	97,319	2,896
5 groups	13,026	25,534	106,783	2,896
100 groups	13,393	25,534	127,699	2,858

Source: Round 1, 2000 MEPS.

Table 2. Percentage distribution of insurance coverage, with standard errors in ( )

NR Adj. Method*	1. all are insured	2. some are insured	3. none is insured
Current method	70.60 (6.57)	12.06 (1.25)	17.35 (7.65)
Direct	70.74 (6.31)	12.18 (1.22)	17.08 (7.36)
5 groups	70.54 (6.44)	12.16 (1.25)	17.03 (7.51)
100 groups	70.30 (6.55)	12.15 (1.27)	17.55 (7.64)

Note: Estimates are for research purposes only and do not reflect national estimates.
Source: Round 1, 2000 MEPS.

Table 3. Percentage distribution of health status, with standard errors in ( )

NR Adj. Method*	1. all are poor/fair	2. some are poor/fair	3. none is poor/fair
Current method	5.72 (0.65)	12.42 (1.24)	81.87 (1.79)
Direct	5.77 (0.64)	12.67 (1.23)	81.57 (1.76)
5 groups	5.70 (0.64)	12.46 (1.23)	81.84 (1.77)
100 groups	5.77 (0.66)	12.39 (1.25)	81.83 (1.80)

Note: Estimates are for research purposes only and do not reflect national estimates.
Source: Round 1, 2000 MEPS.

* NR. Adj. Method = Nonresponse adjustment methods:
      Current method = cell classification as modeled by CHAID and currently used in MEPS.
      Direct = method of using propensity score directly.
      5 groups = method of using propensity scores to classify units into 5 groups.
      100 groups = method of using propensity scores to classify units into 100 groups.

Table 4. Weighted mean and median of health care expenditures

(A) Mean

NR Adj. Method*	Mean	SE of mean	Relative SE of mean
Current method	4811	478	0.0994
Direct	4858	463	0.0953
5 groups	4824	470	0.0974
100 groups	4810	478	0.0994

(B) Median

NR Adj. Method*	Median	SE of median	Relative SE of median
Current method	1675	357	0.2131
Direct	1693	345	0.2038
5 groups	1680	353	0.2101
100 groups	1673	358	0.2140

Note: Estimates are for research purposes only and do not reflect national estimates.
Source: Round 1, 2000 MEPS.

Table. 5. Distribution of DU level Nonresponse Adjusted and Family level Post-stratified Weights, (With response propensity calculated from the logistic model with MEPS base weight included)

NR Adj. Method*	Standard deviation	Mean	Max	Min
Current Method	13,249	25,534	121,915	2,549
Direct	12,614	25,534	97,326	2,895
5 groups	13,301	25,534	107,441	2,895
100 groups	13,714	25,534	149,466	2,819

Source: Round 1, 2000 MEPS.

* NR. Adj. Method = Nonresponse adjustment methods:
     Current method = cell classification as modeled by CHAID and currently used in MEPS.
     Direct = method of using propensity score directly.
     5 groups = method of using propensity scores to classify units into 5 groups.
     100 groups = method of using propensity scores to classify units into 100 groups.