Evaluation of Alternative Propensity Models for Adjusting Weights To Compensate for Dwelling Unit Nonresponse in the Medical Expenditure Panel Survey (MEPS)

Lap-Ming Wun[1], Trena M. Ezzati-Rice[1], Nuria Diaz-Tena[2], and Janet Greenblatt[1]
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850[1]
Mathematica Policy Research, Inc.[2]

Keywords: weighting; response propensity; propensity score; logistic regression; step forward and backward logistic regression; CHAID

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 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, a weighting class adjustment, using the tree algorithm method, the Chi-squared Automatic Interaction Detector (CHAID), is used to form the weighting cells at the dwelling unit (DU), i.e., household, level and to create the nonresponse adjustment cells (Cohen, DiGaetano, and Goksel, 1999). An alternative method is to calculate response propensity from logistic models based on response related covariates to make 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. Simpler versions of this type of adjustment have been studied using earlier panels of MEPS data (Wun et al (2004), and Wun et al (2005a)). With newly identified variables as additional potential covariates for nonresponse adjustment, a more elaborate method of identifying appropriate logistic models for nonresponse adjustment is adopted for the investigation of alternative approach of nonresponse adjustment. In this paper, we report the procedure for building the logistic model for nonresponse adjustment using the 2002 panel 9 MEPS data.

2. 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 2002 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 only reports logistic modeling for 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 the NHIS population estimates to account for NHIS nonresponse and undercoverage. This ratio adjusted base weight is then adjusted for nonresponse in MEPS eligible sample DUs at round 1. More specifically, the base weights of MEPS responding DUs are adjusted to compensate for the nonresponding DUs.

3. 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-

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.

Cerantes (2003)). In our study, 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). Comparisons of the various methods of adjustment using MEPS data for earlier years were reported by Wun et al (2004), and Wun et al (2005a).

4. Nonresponse Adjustment Using Propensity Scores from Logistic Model

A propensity score of response in surveys is essentially the conditional probability of response given the covariates. For our study, it was calculated through the following steps:

1. Identify an appropriate, e.g., the best fit, logistic model from a set of potential covariates with response/nonresponse indicator as the dependent variable
2. Calculate a logit value for each sample unit using the model established in 1 above.
3. Convert the estimated logit value obtained from the logistic model established in step 2 into the predicted probability of response, i.e., the propensity score, using 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:

A. 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.

B. Grouping scores to form adjustment cells:

Using the propensity scores, the sample is grouped into classification cells, then the weights of the respondents are adjusted in the cell to compensate for nonresponse.

In this report, we present and discuss the work of step 1 - identifying appropriate logistic models from potential covariates.

5. Covariates of Response

Since each annual MEPS sample is a subsample of respondents to the previous year's NHIS, survey variables for all the sampled units are available from the NHIS. The following NHIS variables, identified as relevant to response by Cohen and Machlin (1998), with periodic updates, have been 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 – 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 identified (Kashihara, et al (2003)) and added to the list in 2004:

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.

This list of covariates is the potential set of covariates for response. Each year a subset of them are identified in CHAID as significant and used in constructing adjustment cells for nonresponse.

6. Logistic Models

With the 29 potential covariates listed in the last section along with the MEPS base weight as a covariate following the rationale of Little and Vartivarian (2003), we have 30 potential covariates for modeling the response. Seven logistic models were developed and evaluated:

I. Models with main effects only:

1. Inclusion of all 30 covariates
2. Use of step forward method with significance level of 0.3 to select significant terms.
3. Use of step forward method with significance level of 0.1 to select significant terms.
4. Use of step backward method with significance level of 0.3 to select significant terms.

II. Models with main effect and interactions:

5. Unweighted - the MEPS base weight is used as a covariate and an unweighted logistic regression model is identified.
6. Weighted - the MEPS base weight is used as weight in weighted logistic regression.

Models 5 and 6 above kept significant terms but did not eliminate collinearity.

7. Same as model 5 but further eliminated collinearity.

Models 1 to 4 simply use main effects in an unweighted logistic model. In models 5 to 7, second and third order interactions are included in the model besides the main effects. Models 5 and 7 are unweighted, while model 6 is weighted. The approach of selecting the second and third order interactions is the same in models 5 and 6 (based on the CHAID tree branches). Model 7 only uses uncorrelated variables and uses different CHAID trees to see if any branches of the trees detect statistically significant interactions. The result of the first CHAID tree (including all the uncorrelated covariates) indicates interactions conditional to the first node chosen by CHAID. The second CHAID tree included the uncorrelated covariates, except the first node of the first tree and the results of the second tree show other possible interactions conditional to the main effect selected in the second tree. The succeeding trees follow the same pattern; including all the uncorrelated variables except those already selected as first nodes previously. We keep the main effects when an interaction is significant at p-value <- 0.30, we did not include the interactions or main effects if the p-value was larger than 0.30.

7. Evaluation Criteria and Results

In addition to the significance of each of the covariates and interaction terms that were used in identifying each of the seven models, we also looked at various goodness of fit criteria. The four criteria of goodness of fit included: -2LogL, maximum-rescaled R-square, percent of concordance, and p-values of Hosmer and Lemeshow statistics. The results for each model are provided in table 1.

In addition to the goodness of fit criteria, the maximum and minimum relative bias and the sum of the bias squared were computed for each model. The relative bias of each model was calculated using the following formula:

RelBias|model = (true percentage (variable x) - estimated percentage (variable x)|model)/true percentage (variable x)

With information from the NHIS for all the sampled cases for the MEPS, we estimated the "true" weighted percentage of each category for the 29 characteristics known for all the MEPS respondents and nonrespondents. The relative bias calculations were carried out in the following three steps:

1. Compute the "true" percentage for the population under study using the base weights and for all sampled cases (respondents and nonrespondents). For example, we computed that the civilian noninstitutionalized population consisted of: 5.62 percent with elementary school or less education, 10.52 percent that finished elementary school, but did not receive the high school diploma or GED, 27.38 percent that received a high school diploma or GED, 27.64 percent with some college, 16.35 percent with college degree, 8.58 percent with some graduate studies and 3.91 percent unknown educational attainment. We used only the first six categories in our calculations because the last one is constrained to the previous percentages.

2. Compute the same quantity using the direct nonresponse adjustments for each model with only the respondents. For example the estimated population percentages using model 7 and model 2 are: (1) 5.73 percent who study till elementary school or less for model 7 and 5.65 for model 2, (2) 11.00 percent that finished elementary school, but did not received the high school diploma or GED for model 7 and 10.73 for model 2, (3) 26.68 percent that received the High School Diploma or GED for model 7 and 27.34 for model 2, (4) 28.50 percent with some college for model 7 and 27.83 percent for model 2, (5) 16.09 percent with college degree with model 7 and 16.36 percent with model 2, and (6) 8.76 percent with some graduate studies with model 7 and 8.65 with model 2.

3. Compute the relative difference for all the categories unconstrained for the 29 characteristics (covariates of response) previously described for all the models as shown in the above formula.

For purposes of illustration, we computed a relative difference of -4.56 percent with model 7 for the elementary school, but did not receive the high school diploma or GED group, while for model 2 for this category the relative difference was only -2.00 percent.

Table 2 shows the range of the relative bias for each model: the smallest one as the minimum and the largest one as the maximum.

Finally, we computed the sum of the bias square (in percent) using the 29 characteristics for all the non-constrained categories. While the minimum and maximum relative biases emphasize the worse case with lower and upper bounds, the sum of the bias square provides an overall measure of bias for all 29 characteristics. Results are shown in Table 2.

8. Observations and Summary

In table 1, except for the -2LogL, the higher the value of a criterion the better the fit in terms of that criterion. For -2LogL, a smaller value is better. Among the four criteria of goodness of fit, model 1 is the best for three of the criteria: the -2LogL, max-res R-square, and percent concordance. Model 5 (unweighted) has the highest p-value of Hosmer and Lemeshow, and it is second to the best for the -2LogL and max-res R-square. Model 1, as shown in table 2, has the lowest sum of bias square.

With these observations, model 1 is determined to be the best model. That is, the model including all potential covariates is the best. In this case, the most tedious step of identification of significant covariates can be skipped. However, since the differences of the models based on various criteria are not very large, we shall not rule out the second or even third best models. Those models will be further evaluated by applying them to the procedure of adjusting weights to compensate for DU nonresponse and evaluating the impact of alternative nonresponse adjusted weights for selected survey analytical variables. This further evaluation is reported in another paper (Wun, et al (2005b)).

References

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.

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 ASA Proceedings of the Survey Research Methods Section of the American Statistical Association, pp. 3758-3765, (CD-ROM).

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 (2^{nd} 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.

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 ASA Proceedings of the Section on Survey Research Methods, pp. 4625-4631, (CD-ROM).

Wun, L.-M., Ezzati-Rice, T.M., Diaz-Tena, N., and Greenblatt, J. (2005a),"On Modeling Response Propensity for Dwelling Unit (DU) Level Nonresponse Adjustment in the Medical Expenditure Panel Survey (MEPS)," presentation at the Tenth Biennial CDC and ATSDR Symposium

on Statistical Methods, 2005, Washington DC.
Wun, L.-M., Ezzati-Rice, Digaetano, R., Goksel, H., and Hao, H., (2005b),"Using Data from the National Health Interview Survey (NHIS) to Assess the Effectiveness of Nonresponse Adjustment in the Medical Expenditure Panel Survey (MEPS)," to appear on 2005 ASA Proceedings of the Section on Survey Research Methods, (CD-ROM).

Table 1. Goodness of Fit Criteria for 7 Logistic Models

	Models with main effects only				Models with interactions		
	Model 1 All terms	Model 2 Step forward, 0.3	Model 3 Step forward, 0.1	Model 4 Step backward, 0.3	Model 5 Unweighted	Model 6 Weighted	Model 7 No collinearity
-2 Log L	8,334	8,351	8,411	8,365	8,339	8,802	8,449
Max-Res R-square	0.104	0.101	0.091	0.099	0.103	0.100	0.085
Percent concordance	67.7	67.5	66.9	67.3	67.4	67.3	66.5
Hosmer and Lemeshow	0.620	0.052	0.519	0.635	0.800	0.781	0.014

Source of data: 2002 MEPS

Table 2. Bias for 7 Logistic Models

	Percentage Minimum Relative Bias	Percentage Maximum Relative Bias	Sum of Bias Square (percent)
Model 1 All terms	-2.06	1.82	1.10
Model 2 Step forward, 0.3	-2.02	1.68	1.30
Model 3 Step forward, 0.1	-3.14	4.76	3.30
Model 4 Step backward, 0.3	-2.88	2.56	2.92
Model 5 Unweighted	-3.90	3.43	4.44
Model 6 Weighted	-3.71	3.65	4.61
Model 7 No collinearity	-4.56	5.43	6.77

Source of data: 2002 MEPS