

Response Propensity Weighting for the Medical Expenditure Panel Survey - Insurance Component (MEPS - IC)

John Sommers, Agency for Healthcare Research and Quality¹
Steven Riesz, Bureau of the Census²
David Kashihara, Agency for Healthcare Research and Quality¹

Key Words: weighting, health insurance, poststratification, non-response.

Abstract

The MEPS-IC is an establishment survey, sponsored by the Agency for Healthcare Research and Quality (AHRQ), that collects data about employer-provided health insurance. Currently, non-response adjustment and post stratification are done using a cell-based system where cells are defined using employment size classes, states, and industries. This system was developed in the first years the survey was conducted, when no sample data was available. This system and any new system must take into account a complex interview process that creates multiple levels of response and non-response. This paper discusses the evaluation of proposed new methods to adjust the weights for non-response. These new methods use response propensity scores. They allow a larger set of variables to be used in adjusting the weights than the current method. Also, they may produce weights with a smaller range of values. Such methods are likely to reduce both the bias and the variance of the survey estimates. The new methods are evaluated using 2001 survey data, by comparing a variety of estimates and their error estimates to one another, and to those produced by the current method. Recommendations for changes to the current weighting method are made.

Background

The Medical Expenditure Panel Survey - Insurance Component (MEPS-IC) is an annual survey of business establishments (locations) and governments. The survey is funded by AHRQ, and is conducted by the Census Bureau. The first MEPS-IC collected data for the year 1996. Data are collected about various aspects of employer-sponsored health insurance, such as whether health insurance is offered, the number of

employees enrolled in health plans, and the premium amounts, including the employee and employer contributions to the premium.

The survey's goal is to publish quality estimates, for the nation and at least 40 states, of quantities such as the average premiums and contributions per enrollee, the percentage of employees enrolled, and the percentage of establishments and governments that offer health insurance. Because employers are a key source of health insurance in the United States, data are used by federal agencies, such as the Bureau of Economic Analysis, the Centers for Medicare and Medicaid Services, the Department of the Treasury, and by state governments, to monitor and predict national and state trends in employer-sponsored health insurance.

There are two list frames for the MEPS-IC, one for the private sector and one for the public sector. Therefore, two samples are selected:

- A private sector sample of establishments selected from the Census Bureau's Standard Statistical Establishment List (SSEL). (The location of the establishment is important since state estimates are very important, and establishments of multi-location firms can be in different states. A firm is a legal entity that can own one or more establishments, therefore, a sample of firms is a cluster sample that does not consider location.)
- A public sector sample selected from the Census of Governments.

By far the largest portion of the sample is from the

¹ The views expressed in this paper are those of the author 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.

² This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

private sector, where the sample size is approximately 43,000 establishments annually. This paper will focus on improving the non-response adjustment for the private sector sample (Sommers, 2000).

Current Non-Response Adjustment

To understand the non-response adjustment for the MEPS-IC, one must first understand the relationship between the data collection methods and the non-response adjustment method. Data collection is carried out in three stages, and results in five sets of in-scope sample units.

The first stage of data collection is a short telephone interview to determine whether a sampled establishment offers health insurance to its employees. For establishments that do not offer insurance, a short interview is conducted and the case is closed out. This group of respondents is classified as sample group 1. For those that do offer insurance, a mail questionnaire is sent. If this is not returned, a follow-up telephone interview is attempted. Among these units, only those that provide the second set of information are full respondents. This is sample group 2. Those who do not respond to the second stage are placed in group 3 of the sample. These are partial respondents since we know they offer health insurance, which is a key variable in the survey.

Those establishments not contacted during the initial short telephone interview process are contacted using a mail and telephone process. Those that respond at this stage are placed in sample group 4 (these establishments may or may not offer health insurance), and those that do not respond are placed in sample group 5. This latter group is the set of full non-respondents.

Correction for non-response occurs in two steps. First, using a set of cells based upon the employment size of the establishment, the employment size of the firm, the location, and the industry, the sampling weights for respondents in group 2 are adjusted to represent the non-respondents in group 3. This is done because of the partial information shared among these two groups of the sample.

Next, the weights of units in groups 1, 2, and 4 are adjusted to reflect the non-responding units in group 5. The weights adjusted at this stage are the sampling weights for all of the groups except group 2, where the weights that have been adjusted to reflect group 3 are used. After this adjustment, the weights are post stratified to frame employment totals using similar sets of cells.

This weight adjustment method for the current MEPS-IC (Sommers, 2000) is similar to one developed by Westat, Inc., for the 1993 National Employer Health Insurance Survey (NEHIS), which was a precursor to the MEPS-IC (Marker, et. al., 1996). The method uses an iterative procedure that first corrects for non-response using a set of cells defined by industry. After this, these values are adjusted using cells defined by whether the establishment is in a single-unit firm or a multi-unit firm. Lastly, they are adjusted using cells defined by a cross of state, establishment employment size class, and firm employment size class. This process is iterated until the weights converge.

Selection probabilities and thus weights for the MEPS-IC are not equal. They vary considerably, as can be seen in Table A. This is because larger establishments have greater numbers of enrollees than small establishments, and thus have a higher probability of selection. The selection probability of an establishment increases with the employment size of the establishment, but is not directly proportional to the employment size. This was done because although most of the estimates are more greatly influenced by larger establishments, estimates based strictly on the number of establishments are also important. Thus, the sample design must balance the needs of estimates involving the number of establishments, and estimates involving the number of enrollees.

For the 1993 NEHIS, the allocation to each stratum was proportional to the sum of the square root of the establishment employment. This root allocation tended to balance the allocation between the number of establishments and the employment in each stratum (Marker, et. al., 1996). The current allocation for the MEPS-IC is an average of the optimal allocations for several types of estimates (Sommers, 1999).

Since the start of the MEPS-IC, much has been learned about the survey data. For instance, we can see which variables on the frame have a significant impact on the values of the survey variables, and on the likelihood of a unit to respond to the survey. Aside from employment size, industry, and location, there are several other variables on the frame that are highly important in predicting the values of survey variables and response rates. Because all of these frame variables are not currently used in adjusting the weights for non-response, the current method may not be adequately adjusting for non-response, and there may be some bias in the estimates. Thus, a study was performed to find an improved non-response adjustment method.

Proposed New Non-Response Adjustment

The proposed new method for non-response adjustment is based upon the method of response probability or response propensity (Iannacchione, et. al., 1991). This method uses logistic regression models to predict the probability that an individual sample unit will respond. This is known as the response propensity. Non-response adjustments to the original sampling weights of respondents can be performed either (1) by using weighting cells defined by ranges of the predicted response propensity or (2) by dividing the original sampling weight by the predicted response propensity for each respondent.

To apply these methods to the data from the MEPS-IC, two models are required, one to adjust for partial respondents in group 3, and one to adjust for full non-respondents in group 5. The model for partial respondents was determined using only the data from establishments in groups 2 and 3. The model for full non-respondents was determined using the data from all of the establishments, with the establishments in group 5 categorized as non-respondents.

For both models, the independent variables were a set of 9 variables which had proved to be correlates/predictors of several survey variables, such as premium amount, employee contributions to the premium, and employee enrollment rates. These correlates were: firm employment, establishment employment, state, industry, age of firm, whether the firm is a single-unit firm or a multi-unit firm, average wage, type of firm, and type of county (based on an urban/ rural code).

To select the models, both forward and backward selection methods were run in SAS. For the first model, which is used to adjust for partial response, the results of the forward and backward selection methods were the same. The significant variables were firm employment, state, type of firm, industry, type of county, and age of firm. For the model of establishments in groups 1 to 5, the forward and backward selection methods again resulted in the same set of covariates. They were establishment employment, state, industry, age of firm, whether the firm is a single-unit firm or a multi-unit firm, type of county, and average wage.

Once the models were selected, it was necessary to determine how to use them for non-response adjustment. One can create n cells based on the value of the response propensities, where each cell contains $1/n$ of the total sample. Another method is to create the maximum number of cells by adjusting each

respondent's original sampling weight by using its own response propensity directly. We chose two adjustment methods: (1) a 5-cell method, where the cells are defined using the quintiles of the distribution of the propensity scores, and (2) the direct method, where the original sampling weight is divided by the response propensity for each respondent (Little and Rubin, 1987).

The former method adjusts the weights of the respondents so that, in each cell, the sum of the adjusted weights of the respondents equals the sum of the original sampling weights of the respondents and the non-respondents. The latter method simply divides the respondent's original sampling weight by its response propensity. Both adjustment methods are performed in two steps. In the first step, one adjusts the weights of the establishments in group 2 to account for the partial non-response of the establishments in group 3 (using the first model). Then, in the second step, one adjusts the weights of the establishments in groups 1, 2, and 4 to account for the non-response of the establishments in group 5 (using the second model). In this second step, for establishments in group 2, one uses the weights that have been adjusted in the first step.

Analysis of the Effect of the Response Propensity Adjustments

Two methods were used to evaluate the response propensity adjustment methods. The first was to analyze the new weights created, and to compare them to the current weights. This was done by calculating descriptive statistics for the weights before and after non-response adjustments, and comparing these statistics to those for the current non-response adjustment method. The second method of evaluation was to calculate a group of important estimates and their standard errors for two subpopulations, and then to compare these results to the results from the current method.

Table A shows the results of the first evaluation. In this table, we show the mean, median, standard deviation, interquartile range, maximum, and skewness of the weights of the respondents, before non-response adjustment and after non-response adjustment and post stratification. One should note that all three sets of weights are affected by poststratification, and all are changed to about the same extent by post stratification. Also, the change in the weights due to post stratification is much smaller than the change due to non-response adjustment. Thus, the results in Table A reflect differences caused by the non-response adjustments, and not the poststratification.

Table A shows descriptive statistics for respondent weights before application of non-response adjustment and after application of the three adjustment methods. As one would expect, the means, ranges, and maxima increase for all three methods over the original sampling weights of respondents. One would expect an average increase of about 1/(the overall response rate). Because of the individual adjustments to each respondent that occur with the direct response method, one would expect this adjustment to produce weights with the greatest range, largest maximum value, and largest skewness and standard deviation values, and this is what was seen.

An interesting result is that the current method, in spite of its having many more adjustment cells than the 5-cell response propensity method, produced the overall least extreme set of weight values. This method's weights have the smallest range, maximum value, skewness, and standard deviation. Given these differences in the weights and their structures, one would expect that there could be significant differences in the standard errors of the estimates produced, and in the expected values of the estimates produced.

Tables B, C, D and E show results for four important survey estimates: percent of establishments that offer health insurance, percent of employees enrolled, average single premium, and average single employee contribution to the premium. For each estimate, results are shown for two subpopulations- state, and firm size crossed with industry group. The estimates and their relative standard errors under the two new weighting methods were compared to those from the current weighting method. The results are similar for other subpopulations.

There are two items of interest in these tables. First, surprisingly, the averages of the relative standard errors of the estimates are almost equal for each of the weighting methods. This is also true for other subpopulations, for these four estimates. However, when we run 5% significance tests to see if the estimates using the new weighting methods are statistically different from the current estimate, we find that more than 5% of the time, the new estimates are different from the current estimate. These differences indicate to us that although the relative differences are small, the two new weighting methods produce estimates with slightly different expected values than the current weighting method. This is especially true for state-level estimates of the percent of establishments that offer health insurance. This seems reasonable because this estimate is based on only the final weights. Other estimates, such as the percent

enrolled, are based on the weights and other variables, such as the number of enrollees. For these estimates, a unit with a large weight will not influence the estimate very much if its enrollment is small. Units with small weights but large numbers of enrollees will be more influential.

We feel that the propensity score methods should have less bias than the current method. Tables B - E show that there are differences between the estimates using the propensity score methods and the estimates using the current method, and they show that the standard errors are similar for all of the methods. This indicates that the mean squared errors using either of the response propensity methods are on average slightly smaller than the mean squared errors using the current method.

Because we believe the two new response propensity methods are superior to the current method, we compared them to determine which method is superior. Table F shows some typical results. More than 5% of the significance tests showed differences between the estimates using the two methods. However, the mean relative differences between the estimates were very small, so there is little practical difference between the methods.

Conclusions

We compared a variety of estimates produced using the current MEPS-IC non-response weight adjustment method with two proposed response propensity weight adjustment methods. All three methods result in very similar standard errors. However, there is a small but significant difference between the expected values of the estimates using the current method and the estimates using the proposed new methods. In examining the results produced by the two new methods, little practical difference was found. Thus, either of the two new methods should be implemented.

Future Work

We intend to do a small amount of extra research on the creation of the cells for the response propensity adjustment method. The current 5-cell method is based upon the quintiles of the distribution of the response propensity scores. However, because this distribution is rather skewed, we are considering a different method of creating boundaries that would work in a similar fashion to the methods of producing stratum boundaries for sampling. After this work, we will pick a final method to implement. If we cannot find any practical differences between the various new methods, then we will choose the new method that is

the easiest to implement.

References

Cochran, WG. Sampling techniques. New York: John Wiley and Sons: 1977.

Iannacchione, JG, Folsom, RE. "Response Probability Adjustments Using Logistic Regression." 1991 Proceedings of the Survey Research Methods Section of the American Statistical Association, 637-642.

Little, RJA, Rubin, D. Statistical Analysis with Missing Data. New York: John Wiley and Sons: 1987.

Marker D, Bryant E, Wallace, L, Yansaneh I. National Employer Health Insurance Survey (NEHIS): Volume I; statistical methodology. Rockville(MD); Westat, Inc.; 1996.

Sommers, JP. List sample design of the 1996 Medical Expenditure Panel Survey Insurance Component. Rockville (MD): Agency for Health Care Policy and Research; 1999. MEPS Methodology Report No. 6. AHCPR Pub. No. 99-0037

Sommers, JP. Construction of weights for the 1996 Medical Expenditure Panel Survey Insurance Component list sample. Rockville(MD); Agency for Health Care Policy and Research; 2000, MEPS Methodology Report No. 8, AHCPR Pub. No. 00-0005.

Table A

Descriptive Statistics of Weights Before and After Non-response Adjustment and Poststratification							
	Mean	Std Err	Median	Inter-quartile range	Max	Min	Skewness
Before Adjustment (i.e. original sampling weights)	161.64	161.89	118.03	210.72	2847.9	2.0	3.30
After Current Adjustment	211.16	225.62	143.86	262.07	4190.9	0.221	3.52
After Response Propensity Cell Adjustment	212.25	233.90	143.30	260.89	4575.2	0.203	3.95
After Response Propensity Direct Adjustment	212.36	235.3	143.95	261.54	4695.8	0.203	4.19

Table B

Comparisons of Results of New Weighting Methods to Results from Current Weighting Method: Percent of Establishments that Offer Health Insurance							
WEIGHTING METHOD							
	Current Weights	5 Cell Response Propensity			Direct Response Propensity		
By Variables	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE
States	0.0484	-0.00531	45.1	0.0490	-0.00557	51.0	0.0490
Size x Industry	0.0203	-0.00231	18.5	0.0208	-0.00286	25.9	0.0208

Table C

Comparisons of Results of New Weighting Methods to Results from Current Weighting Method: Percent of Employees Enrolled							
WEIGHTING METHOD							
	Current Weights	5 Cell Response Propensity			Direct Response Propensity		
By Variables	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE
States	0.0517	0.00154	17.7	0.0514	0.00115	21.6	0.0515
Size x Industry	0.0415	0.00062	18.5	0.0412	0.00113	29.6	0.0411

Table D

Comparisons of Results of New Weighting Methods to Results from Current Weighting Method: Mean Employee Single Contribution							
WEIGHTING METHOD							
	Current Weights	5 Cell Response Propensity			Direct Response Propensity		
By Variables	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE
States	0.1048	-0.00020	17.7	0.1047	0.00000	13.7	0.1045
Size x Industry	0.1102	0.00030	18.5	0.1112	-0.00070	14.8	0.1117

Table E

Comparisons of Results of New Weighting Methods to Results from Current Weighting Method: Mean Single Premium							
WEIGHTING METHOD							
	Current Weights	5 Cell Response Propensity			Direct Response Propensity		
By Variables	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE	Mean Relative Diff	Percent Significantly Different from Current	Mean RSE
States	0.0363	-0.00033	13.7	0.0365	-0.00089	9.8	0.0365
Size x Industry	0.0330	0.00004	14.8	0.0338	-0.00078	11.1	0.0335

Table F

Comparisons of Results of Two New Weighting Methods					
Estimate Type	By Variables	Mean Relative Difference	Percent Significantly Different	Mean RSE 5 Cell Method	Mean RSE Direct Method
Percent of Establishments that Offer Insurance	States	0.00027	35.3	0.0489	0.0490
	Size x Industry	0.00055	7.4	0.0207	0.0208
Percent Enrolled	States	0.00039	19.6	0.0515	0.0515
	Size x Industry	-0.00050	18.5	0.0412	0.0411
Mean Single Employee Contribution	States	-0.00020	7.8	0.1046	0.1045
	Size x Industry	0.00100	14.8	0.1111	0.1117
Mean Single Premium	States	0.00056	11.8	0.0365	0.0365
	Size x Industry	0.00082	11.1	0.0337	0.0335