A New Stratification for the Medical Expenditure Panel Survey - Insurance Component (MEPS - IC)

John Sommers, Agency for Healthcare Research and Quality[1]
Steven Riesz, Bureau of the Census[2]

Key Words: MEPS-IC, stratification, employer-based sample, employer-sponsored health insurance

Abstract

The MEPS-IC is an establishment survey, sponsored by the Agency for Healthcare Research and Quality, that collects data about employer-provided health insurance. Currently, stratification is based upon the employment sizes of the establishment and the controlling firm. This paper discusses the evaluation of a proposed new two-way stratification scheme. The new strata are defined, and their boundaries are determined, by using predicted values of two important survey variables, the probability that an establishment offers health insurance and the number of enrollees at an establishment. These variables have very different sampling requirements. Generally, the first requires a large sample of small establishments, while the second requires a large sample of large establishments. Several allocations are evaluated using variance components estimated with past survey data. The results are compared to the current stratification and allocation method. Recommendations for changes to the sample design 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 the Agency for Healthcare Research and Quality (AHRQ) and conducted by the Census Bureau. The first survey collected data for the year 1996. Data are collected on 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 contribution to the premium.

The survey's goal is to publish quality estimates, for the nation and 40 states, of quantities such as the average premium and contribution 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, and 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 survey, 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 private sector where the sample size is approximately 38,000 establishments annually. This paper will focus on improving the stratification and allocation of the private sector sample (Sommers, 1999).

Current Stratification and Allocation

In order to understand the stratification and allocation of the MEPS-IC, one must first understand the estimates that must be made. They fall into two categories:

[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 paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

- Estimates related to the number of establishments, such as the percentage of establishments that offer health insurance.

- Estimates related to the number of employees, such as the number of employees enrolled in health insurance plans, the total cost of premiums for all enrolled employees, and the total cost of employee or employer contributions to premiums for all enrolled employees.

These two types of variables require allocations that are almost the exact opposite of one another. Most establishments are in small firms that are less likely to offer health insurance and have only a small portion of the total enrollment. On the other hand, the larger establishments from the largest firms are almost certain to offer health insurance and have a large portion of the total enrollment. Thus, a sample that is optimal for the first type of estimate would concentrate heavily on small establishments in small firms, while a sample that is optimal for the second type of estimate would focus on large establishments from large firms. A sample that is able to support both types of estimates is a compromise between the two optimal samples.

The present stratification (Sommers, 1999) is the same as one developed by Westat, Inc. for the National Employer Health Insurance Survey (NEHIS) conducted in 1993 for several agencies within the Department of Health and Human Services, including AHRQ. This survey was a precursor to the current MEPS-IC. Similar information was collected and similar types of estimates were required (Marker, et. al., 1996). Within each state there are 14 strata. Strata are defined on two dimensions, employment size of the controlling firm, which relates best to the chance that a location offers health insurance, and employment size of the establishment, which correlates well with the number of enrollees and other variables. There are 3 establishment employment size strata for small firms, 5 for medium-sized firms, and 6 for large firms. There are fewer strata for establishments of small firms because the size of the firm limits the size of its establishments, for example, a firm with 10 employees cannot have an establishment with more than 10 employees. However, a firm with 10,000 employees can have establishments of many sizes, so more strata are required for large firms.

For the 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).

For the 1996 MEPS-IC, the allocation was based upon the weighted average of strata variance components that were estimated using the NEHIS data. Since that time the allocation has been adjusted slightly to an average of the optimal allocation for estimating the total number of establishments that offer health insurance and the optimal allocation for estimating the total number of enrollees. (The optimal allocations were determined by using variance components that were estimated with NEHIS data.)

Proposed New Stratification

The development of new strata was motivated by work on models that were being tested for use in small area estimation with MEPS-IC data. It was realized that although employment size is related to the chance that an establishment offers health insurance and the number of enrollees, other independent variables that are available on the frame are also correlated with the outcome variables of interest. Among these independent variables are average wage, age of the firm, state, and industry. From this work, it was decided that new strata could be defined based on two predicted values- the predicted probability that an establishment offers health insurance and the predicted number of enrollees at an establishment. Since strata should be defined using variables that correlate best with the outcome variables of interest, these predictions, which are based on several variables that are correlated with the outcome variables of interest, could be used to define strata that might have less variation in the important outcome variables than strata that are defined using employment size alone.

Using the two predicted values, a new set of 14 strata was defined. As with the old stratification method, a two-way stratification was created. The first step used the cumulative square root of f rule to create 6 strata based on the predicted probability of offering insurance. After this was done, the 3 strata that contained establishments with the highest probabilities of offering health insurance were each broken into substrata using the cumulative square root of f rule applied to the predicted enrollment at an establishment. The stratum containing establishments with the highest probability of offering insurance was broken into 6 substrata, the next highest was broken into 3 substrata, and the third highest was broken into 2 substrata.

The reason for the decreasing number of substrata as the probability that an establishment offers health insurance decreases is that almost all large establishments offer health insurance. Thus, the largest establishments are in

the stratum with the highest probability of offering health insurance. Only this stratum and the next two strata containing establishments with the greatest likelihood of offering health insurance need to be broken apart by predicted enrollment size. The three strata of establishments with the lowest probability of offering health insurance contain only small establishments. The establishments in these strata have an average size of less than 5 employees. Since there is little variation in the size of these establishments, there is no reason to break them apart into further substrata based on enrollment size.

Analysis of the Effect of the Proposed New Stratification

In order to analyze the effect of the proposed new stratification, variance components for the 14 old strata and the 14 proposed new strata were estimated for 2 variables, whether or not an establishment offers health insurance and the enrollment for an establishment. The variance of a total using a stratified sample with simple random sampling without replacement in each stratum is

$$\sum_h \frac{N_h^2}{n_h}\left(1-\frac{n_h}{N_h}\right)S_h^2,$$

where N_h is the size of stratum h, n_h is the sample size for stratum h, and S_h^2 is the population variance within stratum h (Cochran, 1977). Thus, in order to estimate this variance, we need to estimate S_h^2. To do this, we used the following procedure. First, we compute

$$\hat{S}_{hij}^2 = \frac{\sum_k finalwgt_{hijk}(y_{hijk}-\bar{y}_{hij})^2}{\sum_k finalwgt_{hijk}},$$

where h represents strata within states, i represents the year of the survey data, j represents states, k represents establishments, y is the survey variable of interest, and finalwgt is the sampling weight after it has been adjusted for unit nonresponse and poststratified. $\hat{S}_{hij}^2$ is an estimate of S^2 for stratum h, in year i and state j. We combine these $\hat{S}_{hij}^2$'s in the following manner in order to get a good estimate of S_h^2, the population variance for stratum h:

$$\hat{S}_h^2 = \frac{\sum_{ij} w_{hij}\hat{S}_{hij}^2}{\sum_{ij} w_{hij}},$$

where $w_{hij} = \sum_k finalwgt_{hijk}$.

Assuming that a typical state has a distribution of establishments similar to that of the entire country, variance estimates could be made for any allocation using stratum sizes which are available from the frame and values of the variance components which were estimated using three years of MEPS-IC data.

The standard errors of two totals were estimated for sample designs using the old or new strata with the following allocations: (1) optimal for the total number of establishments that offer health insurance, (2) optimal for the total number of enrollees, (3) proportional to the square root of the sum of establishment employment, and (4) the average of the two optimal allocations. An estimate was also made using the old strata with the current allocation. The optimal allocations were used to determine the best that could be done for each variable, the root and average allocations were used because they have been used in the past, and the current allocation was used to provide a benchmark. The estimates of the standard errors are shown in Table A.

As one can see from Table A, the new stratification had a very significant effect on the error for total enrollment, causing a drop in the optimal error of 53.8 percent. The drop in the optimal error for the number of establishments that offer health insurance was 7.0 percent. The average of the two optimal allocations, which allows for a balanced allocation that yields good estimates for both totals, yields similar results when compared to the average of the two optimals using the old stratification. The current allocation, which is based upon an average of optimal allocations using variance components that were estimated with 1993 NEHIS data, seems to have held up well. Under the old strata, the standard errors produced by the current allocation are very close to those produced by the average of the optimal allocations that were obtained from more recent variance component estimates.

Further Analysis of the New Stratification

Given the results in Table A, we decided to examine the

new stratification further, including new allocation methods. The assumed sample size was the average sample size of the last three survey years. As part of this process, another important survey variable was added to the analysis- the total cost of employee contributions to the single premium at an establishment. This variable, which is closely related to enrollment, was added because we decided that more weight should be given to enrollment-related variables. We also wanted to check that the new stratification would have an effect on errors of estimates that involve enrollment-related variables similar to the effect that it had on the estimate of total enrollment.

Several allocation methods were tried. In addition to the root allocation, we tried a weighted average of the optimal allocations for the three estimates, and we applied a Neyman type allocation to variance components which were weighted averages of the variance components of the three analysis variables (Cochran, 1977). The goal was to find an allocation in which the increase in the standard errors of the three estimates over their optimal standard errors was approximately equal. This balance among the estimates involving the three variables is necessary since the best allocation for each variable is a very poor allocation for at least one of the other two variables. This can be seen in Table A by comparing the standard error of an estimate under its optimal allocation to its standard error under the other estimate's optimal allocation.

As various allocations were tried, we realized that there was a very small percentage of the establishments that should be designated as certainties. These establishments had very large predicted enrollments. Designating these large establishments as certainties reduced the size of the variance component for the stratum containing establishments with the largest enrollment. This reduced the optimal standard error by approximately 10 percent for both the estimate of total enrollment and the estimate of total single enrollee contribution. It had no effect on the optimal error for the estimate of the total number of establishments that offer health insurance.

As was our experience when the original allocation method was selected (Sommers, 1999), several allocations yield similar standard errors for each variable. The one that was chosen gave good results for all three variables, and was easier to implement than some of the other choices. The method uses a Neyman type allocation based upon variance components which are weighted averages of the variance components for the three variables being considered. The last two rows of Table B show standard error estimates for this proposed weighted allocation using the new strata, and for the current allocation using the old strata. The projected decreases in the standard errors of the three estimates using the proposed weighted allocation and the new strata, compared to the current allocation and the old strata, are 12.7%, 57.5%, and 49.3%, respectively.

The first three rows of Table B show standard errors for the optimal allocations for three different stratifications. Note that these first three rows show the optimal values obtained using three different allocations, and therefore each row does not correspond to one sample design- they only show the result of the best allocation for each variable. Also in Table B, one can see the effect of the certainty stratum on the standard errors for the estimates of total enrollees and total single employee contributions.

Conclusions

We have examined the use of strata defined by predicted values of two important survey variables. These predictions are based on several variables which are known for each establishment on the frame, and which have correlated well each year with important survey data. Using the two predictions, we have proposed a new stratification. Using variance components that were estimated using multiple years of survey data, we have shown that the new stratification shows promise for producing important estimates with standard errors that are significantly less than those the current simpler stratification produces. These current strata are only based upon two variables which are correlated with the survey data rather than the much larger number of variables used in the prediction models. Using the new stratification could especially increase the overall quality of state-level estimates, where the current small sample sizes only allow for estimates of useable but marginal quality.

Future Work

Our next goal is to review our sampling, weight adjustment, and estimation procedures to determine changes that would be needed to implement the new stratification process, and then to carry out the implementation, if possible. After this, we intend to analyze the real effect of the new stratification on the errors of important state and national estimates. If the errors are significantly smaller, then we intend to reconsider our method of allocating the sample to the states. By changing this method, we hope to be able to increase the sample sizes sufficiently in the smallest states, whose current estimates have errors large enough

to prevent their publication. At the same time, we will decrease the sample size somewhat in states which currently have publishable estimates. Our overall goal is to publish estimates for all states, without decreasing the quality of our currently published national and state estimates.

References

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

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.

TABLE A

Standard Errors for National Trial Allocations Using Old and New Strata				
	Old Strata		New Strata	
Totals Estimated / Allocation	Establishments Offering Health Insurance	Number of Enrollees	Establishments Offering Health Insurance	Number of Enrollees
Optimal for Number of Establishments Offering Health Insurance	15,088	4,626,353	14,025	5,054,792
Optimal for Number of Enrollees	25,052	840,347	27,173	388,512
Root	19,862	1,047,248	17,885	772,031
Average of the Optimals	17,827	997,305	16,673	471,909
Current	18,802	978,356	Inapplicable	Inapplicable

TABLE B

Standard Errors for Allocations Using Old and New Strata			
Totals Estimated / Strata, Allocation Method	Establishments Offering Health Insurance	Number of Enrollees	Single Employee Contribution
New Strata without Certainties, Optimal	14,025	388,512	1.678×10^8
New Strata with Certainties, Optimal	14,025	354,813	1.606×10^8
Old Strata, Optimal	15,088	840,847	2.746×10^8
New Strata with Certainties, Proposed Weighted Allocation	16,419	415,652	1.793×10^8
Old Strata, Current Allocation	18,802	978,346	3.535×10^8