

SUPPORTING STATEMENT

Part B

**2010 -2011
Medical Expenditure Panel Survey –
Insurance Component**

August 17, 2009

Agency for Healthcare Research and Quality (AHRQ)

Table of contents

B. Collections of Information Employing Statistical Methods.....

- 1. Respondent Universe and Sampling Methods.....3
- 2. Information Collection Procedures.....8
- 3. Methods to Maximize Response Rates11
- 4. Tests of Procedures11
- 5. Statistical Consultants12

B. Collections of Information Employing Statistical Methods

1. Respondent Universe and Sampling Methods

The total budgeted sample size is approximately 45,000 sample units before non-response and out-of-scope units. The sample includes private sector establishments which employ at least one person, plus State and local governments. The sampling goal is to produce adequate estimates for 1) the private sector for all 50 States and the District of Columbia; 2) State and local governments by Census Division; and 3) the Nation as a whole. This goal is discussed later in this section.

The sample frame is derived from two lists:

\$ The Census Bureau's Business Register (BR), a list that contains private sector establishments in the United States which employ at least one person. The list is derived from tax records, and is continually updated to add newly created establishments (births) and remove those establishments that have closed. This list contains over 7,000,000 establishments and is very complete, with any under-coverage caused by cases of domestic workers not reported for tax purposes.

\$ The Census of Governments, which is collected every five years with data updated in non-Census years using a sample survey. Currently, the most recently available Census of Governments is for the year 2007. The Census of Governments contains over 90,000 units.

Together these two lists cover almost 100 percent of all organizations with at least one employee in the economy, excluding the Federal government.

In order to meet the goals of the survey, the sample is developed in several steps. These are:

- \$ the sample is allocated to the state and local government and private sectors
- \$ a small sample is set aside for certainty units in the private sector
- \$ the remaining sample within the private sector is allocated by state and the government sample is allocated by Census Division
- \$ the sample is allocated to strata within each sector.

Allocation to the State and Local Government and Private Sectors

The division of sample between the state and local government and private sectors is based upon past allocations. There are several precision goals for the survey. There are National and State goals for the private sector, and National and Census Division goals for government.

Among the national relative standard error (RSE) goals for the private sector for the survey are the following:

- a .005 RSE for national estimates of single and family premiums
- a .0150 RSE for national estimates of single and family employee contributions
- a .0075 RSE for national estimates of important proportions, such as the percent of employees enrolled in health insurance

Among the national RSE goals for the state and local government sector for the survey are the following:

- a .0075 RSE for national estimates of single and family premiums
- a .020 RSE for national estimates of single and family contributions
- a .010 RSE for national estimates of important proportions

State estimate goals for the private sector are that the RSE for state estimates have errors less than 6 times the similar national private sector goals. Census Division goals for the state and local government sector are that the RSE be less than 5 times the national state and local government goals.

These goals reflect the preference among users for higher quality private sector estimates as opposed to estimates for governments.

Given these goals and the budget limitation for the sample of about 45,000 units, approximately 41,850 sample units are allocated to the private sector and 3,150 to governments. This allocation was based upon the RSEs from past surveys.

Within the private sector, the allocation is broken into 2 parts:

- a small number of approximately 100 large certainty units,
- the remaining sample which is allocated to individual states.

Under earlier clearances, a birth sample was also included in the allocation. But the switch to collecting the data during the reference year (rather than the following year) eliminated the need for a birth sample.

State Allocation for the Private Sector

The optimal national allocation to states would be to allocate proportional to the size of each State. However, for most states this would result in far too small a sample to meet state estimation goals. The 9 largest states under proportional allocation would receive over half the sample and many of the remaining states would have a sample of less than 100 establishments. From experience with past MEPS-IC surveys, it has been determined that a sample of approximately 500 establishments per state yields estimates that meet most state estimation goals using state stratification and allocation processes. To meet state precision goals, an equal size sample could be allocated to each state. An

allocation of equal sample to each state would produce state estimates that meet state estimation goals, but would be 50% less precise nationally than proportional allocation and would not produce adequate national estimates.

The compromise allocation starts by proportionally allocating 17,000 responding sample establishments among the states. The allocation is then supplemented for the 42 smallest states so that each of the 11 smallest states receive about 495 sample establishments and each of the next 31 largest states receive 535 sample units. The 9 largest states receive their entire sample allocation from the proportional allocation of the 17,000 units. This allocation has an error about 20% higher than if the entire sample were proportionally allocated. However, national estimation goals should still be met and the state goals should also be met, with the exception of the 11 smallest states where target quality could be missed by a small degree.

Once the desired responding sample size is determined, the sample allocations are increased to allow for expected nonresponse and out of scope establishments to arrive at the final sample size per state. These sizes are listed in Attachment D.

Private Sector Sample Selection

Once the allocations are complete (excluding the certainty units) for each state, samples are selected for the private sector within each state. The frame to be used for the private sector will be the most recent update of the Business Register (BR). The BR currently contains about 7.0 million establishments. Before final sample allocation and selections, the universe in each state is stratified into cells. The number of stratification cells is different for smallest states versus all other states. There are 8 stratification cells for small states and 14 stratification cells for all other states.

In addition, there is a certainty stratum for each state which contains establishments with projected enrollments of above 5,000 employees. The sample for certainty establishments is allocated for the country as a whole and is not part of the state allocation process.

The table in Attachment E gives the breakdown for small states and for all other states of the percent of private sector establishments in each strata and defines the strata boundaries for survey year 2008. (Note that the strata are defined as a combination of establishment size and firm size, which is different than the definition used in the previous clearance. The current strata design is actually a reversion to the strata design used in earlier years, which proved to be more efficient.)

Once these cells are created, the frame within each state is broken into 8 or 14 strata depending on the state size. The breakdown is made by state to allow for the best sample within each state to assure quality state estimates. The beginning allocation to an individual cell within a state is:

$$n_{si} = \frac{N_{si} \sigma_{si}}{\sum_{k=1}^{14} N_{sk} \sigma_{sk}} n_s$$

where

N_{si} is the number of establishments in the i th stratum in the s th state,

σ_{si}^2 is a unit variance for the s th state and the i th stratum determined using 2008 MEPS-IC data, and

n_{si} is the allocation to the i th stratum in the s th state.

The unit variance is an average of unit variances of a selection of important variables derived from past data. One variance is used for similar cells across all small states; all other states have their own variance that is used across similar cells. These were derived by pooling data across all states for the same strata. This was done because an examination of state by state variances failed to provide significant evidence that the unit variances differed across all small states or across all other states.

The allocation is not an optimal allocation for any one variable. There is interest in a wide array of estimates. Some are establishment population estimates, such as the percent of all establishments which offer health insurance. These types of variables require large samples of small establishments. Others are employee population estimates, such as total enrollments in health insurance, which require the sample to have relatively large numbers of establishments with many employees. The final allocation is the weighted allocation obtained by taking the weighted value of the optimal allocations for several variables and provides a compromise allocation. Trial allocations using this variance compare favorably with the optimal allocations for the individual variables. Results are monitored annually and efforts are continually made to improve strata definitions and allocation to strata.

An allocation to the strata based upon proportions of establishments in the states is given in Attachment E. This gives a reasonable approximation to the proportion of the private sector sample allocated in each stratum within a state, broken out for small states and all other states separately. (These allocations are reviewed each year for potential non-response and out-of-business. Allocations are increased as necessary within each cell to reflect these losses.)

Once these allocations are done each year, they are checked. After these checks are performed, each establishment in a cell is given the same chance of selection equal to

$p_{si} = a_{si}/N_{si}$ where a_{si} is the final cell size allocation within the state.

At this point, in order to create a more efficient sample and to reduce burden on large

firms -- where a single respondent may sometimes provide the information for more than one establishment owned by that firm -- these probabilities are altered for some establishments.

The values of the p_{si} 's for all establishments linked to the same firm on the BR are summed. This yields the number of establishments that are expected to be selected for that firm. For a small number of firms, this expected value is too large and must be reduced to minimize burden. This reduction is also beneficial for the sample, since the insurance offered in establishments with a common owner is correlated.

To reduce this expected number of establishments, the probabilities of selection are simply reduced to a level where their sum is the desired new sample for that firm. To make up for this reduction in sample, the probability of selection for all other establishments in a cell which contains an establishment with a reduced probability of selection is increased by the amount necessary to have the sum of the probabilities of selection within the strata equal a_{si} .

Once these probabilities of selection are finalized, the allocated samples are selected using sequential sampling with a random start – separate sorts are done for small states and for all other states. To perform this selection, the file is sorted by state, strata, industry and employment size. This assures a good balance of establishments within strata.

Allocation and Selection of the State and Local Government Sample

The state and local government sample will use the 2007 Census of Governments as the frame. This contains approximately 90,000 governments. For this selection there will be only two strata per Census Division. There is a certainty stratum which includes all governments with over 5,000 employees. The certainty stratum for governments accounts for approximately half of all government employment.

The non-certainty governments stratum contains all other governments. A sample size of 180 governments is allocated to the non-certainty government stratum for each of the 9 Census Divisions.

To perform the selection, each non-certainty government is given a measure of size equal to the square root of its total employment. This increases the sample of smaller governments relative to their total employment. The selection probability for a single government is determined as the total final Census Division non-certainty state government allocation, times the government's measure of size, divided by the sum of all measures of size within the Census Division.

The non-certainty government sample within each Census Division is selected sequentially from a file sorted by state, type of government (county, city, school district, etc.), and employment using a random start.

Sampling of Health Insurance Plans

In private sector and non-certainty government cases where an employer offers several health insurance plans, a procedure for scientifically selecting a sample of plans will be implemented. The first step is to identify these cases. During telephone prescreening, establishments will be asked to provide the number of health insurance plans they offer their employees. If there are four plans or less, all are taken with certainty. For the few establishments that offer more than four plans, the names of the plans are collected along with estimates of enrollments. The CATI software will be used to randomly select a subset of four plans for collection. The respondent will receive a preprinted plan collection supplement for the selected plans.

All plans are collected for large certainty governments.

Response Rates

The following table presents the approximate sizes of each of the universes from which sample data are being collected, the approximate expected sample size including out-of-scope, the expected in-scope sample size, the expected number of responses, and the response rates. The response rates are based upon experience from past MEPS-IC data collections.

Projected Response Rates: 2010-2011 MEPS-IC

Response Group	Universe	Sample	In-Scope	Respondents	Response Rate
Private Sector	7.0 x 10 ⁶	41827	39,142	31,383	.802
State & Local Government	90,000	3,130	3,026	2,940	.972
Total		44,957	42,168	34,323	.814

2. Information Collection Procedures

Data Collection

Data collection for the MEPS-IC takes place in three phases: prescreening interview, questionnaire mailout, and nonresponse follow-up. The prescreening interview is conducted by telephone. Its goal is to obtain the name and title of an appropriate person in each establishment to whom a MEPS-IC questionnaire will be mailed. Interviewers also verify addresses and identify businesses that no longer exist, have closed, have merged, etc. Establishments from certain large private firms and state and local governments are not prescreened because they are known to offer health insurance. Due

to their size and importance to all Census establishment surveys, the Census Bureau maintains up-to-date contacts with these employers.

For establishments which do not offer health insurance, a brief set of questions about establishment characteristics is administered at the end of the prescreening interview to close out the case. The MEPS-IC questionnaires are mailed to those establishments which, during the prescreening phase:

- \$ were not contacted,
- \$ refused to cooperate,
- \$ were contacted and acknowledged that they did provide health insurance,
- \$ were from large firms or governments specified at the start of collection for mail-only, or
- \$ had no known phone number.

Establishments which do not respond to the initial MEPS-IC mail questionnaire are mailed a nonresponse follow-up package. Those establishments which fail to respond to the second mailing are contacted for a telephone follow-up using computer-assisted interviewing.

Data for the largest state and local governments and private sector firms, reporting for multiple establishments, are collected using specialized staff and forms. This is done to make the collection process flexible, simple, and as little burden for these important respondents as possible. Sometimes multiple telephone contacts and person visits are used to collect these data. For some collections, abstraction from company records or plan brochures is used if the firm insists on such methods.

Samples of survey materials for the MEPS-IC are included in Attachments F-P. Included are all letters, questionnaires and the Computer Assisted Telephone Interviewing (CATI) text used for the prescreener:

- \$ Letter for original mailing to a private sector sample unit, Attachment F
- \$ Letter for original mailing to a government sample unit, Attachment G
- \$ Letter for follow-up mailing to a private sector sample unit, Attachment H
- \$ Letter for follow-up mailing to a government sample unit, Attachment I
- \$ Letter for follow-up fax to a private sector sample unit, Attachment J
- \$ Letter for follow-up fax to a government sample unit, Attachment K
- \$ Thank you letter, Attachment L
- \$ Questionnaire 20(D) – Definitions, Attachment M

- \$ Text of CATI prescreener, Attachment N
- \$ Establishment-level questionnaires, Attachment O
- \$ Plan-level questionnaires, Attachment P

Materials displayed are the most recent version of the respective item. Some minor changes will be made in the final materials to reflect current needs. For example, signatures on letters are updated to reflect current management at AHRQ and Census.

Weighting

The private sector and state and local government samples are weighted separately. Beginning with the inverse probability of selection as a base weight, the data are adjusted for non-response. Logistic regression is used to determine variables which relate to response. Among other characteristics, the number of plans, size of establishment, size of firm or government unit, industry, and whether the respondent received a mail survey must be considered. Adjustments are made so that the sum of respondents in a cell equals the beginning weighted total for the cells. It would be desirable to adjust for non-response in each state, but cell size may preclude that adjustment in all but the largest states.

Once adjustments for non-response are made, data are post stratified to a set of new control totals. For the private sector, this would be new frame counts from the most recent BR for cells determined by state, industry, size of firm and size of establishment. For governments, post stratification is done to counts provided by the Census of Governments Division. Cells are determined by size of government and state.

Estimation and Accuracy

Estimation is done using sampling weights and variances are calculated using standardized software available, such as SUDAAN or the random group method. Both account for the specialized sampling methods used for the survey.

Certain key variables, such as premiums, contributions and enrollments are imputed when item non-response occurs and values are missing. Selection of donors is accomplished using a nearest-neighbor type hot-deck process which chooses the best donor given a set of matching variables and their order of importance. Variables used to match to determine donors are chosen for their correlation with the variable to be imputed, with special care taken to select variables which are also correlated with non-response.

Actual values are derived from donors in various ways. In some cases, ratios are derived from donors and applied to other values the recipient has reported. Great care is taken to maintain consistency of relationships within the data by using ratios and other means that

apply data from the donor to data which have been reported by the recipient. The data are also processed in a specific order to assure that variables important to the imputation of variables later in the process are imputed first. For instance, if one must impute the total single premium and total family premium for the same health plan, the values are imputed in order so that the imputed family premium will depend on the single premium that has already been imputed. This is necessary because these values are highly correlated.

3. Methods to Maximize Response Rates

To achieve maximum response rates the following methods are to be used:

- 1) Perform a screening phone call to identify the best contacts, number of insurance plans and to complete simple cases, such as establishments which offer no insurance.
- 2) Interviewers schedule calls at convenient times for respondents.
- 3) Supervisors regularly attempt to convert refusals.
- 4) Mail a self-administered questionnaire.
- 5) Mail a second questionnaire to those not responding in adequate time.
- 6) After an adequate interval, do a telephone follow-up to either remind the respondent or to collect the data.
- 7) For questionnaires returned by mail which fail key edits, perform a callback to verify data.

A further method to improve response, developed to overcome problems encountered in a 1994 predecessor survey and the 1996 MEPS-IC, is a special handling group. This group consists of highly trained analysts, interviewers and statisticians. Their purpose is to collect data from large firms with high burdens. These respondents are given a primary focus early in the data collection process. Each large firm is assigned a collection coordinator. The special handling group can use any reasonable means to accommodate respondents. They can use personal visits and arrange to collect information in formats easiest for the respondents. There are also special collection forms that allow these firms to list plans only once that may be repeated across multiple establishments. Improved methods of collection for large firms allowed the Census Bureau to significantly improve response rates after the initial 1996 MEPS-IC survey. Another positive effect of these efforts has been a lowering of burden for the larger firms and large governments.

4. Tests of Procedures

This is a renewal of a current survey. As part of general data collection activities, data collection results, interviews, comments made by respondents and estimates are

monitored. New data items added to collection are pretested under a separate Census Bureau testing clearance. Because of this monitoring and collection experience, no special pretest is required at this time for the general survey questionnaire.

5. Statistical Consultants

Dr. John Sommers - RETIRED
Mathematical Statistician
Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality
Rockville, MD 20850

Michael Kornbau
Chief, Program Research and Methods Branch
Bureau of the Census
Washington, DC 20233
301-763-6779

Anne Kearney
Program Research and Methods Branch
Bureau of the Census
Washington, DC 20233
301-763-6780

David Kashihara
Mathematical Statistician
Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality
Rockville, MD 20850
301-427-1474

Dr. Steven B. Cohen
Center for Financing, Access and Cost Trends
Agency for Healthcare Research and Quality
Rockville, MD 20850
301-427-1406