MEPS

Methodology Report #17

Additional Imputations of Employer Information for the Insurance Component of the Medical Expenditure Panel Survey since 1996

ABSTRACT

This report describes the process used to impute values for missing establishment and plan characteristics for the Insurance Component of the Medical Expenditure Panel Survey that are currently imputed but were not imputed for the 1996 Insurance Component. (The list of items imputed for the 1996 Insurance Component, and which continue to be imputed, can be found in MEPS Methodology Report No. 10.) The process involves four types of cases: list sample, private sector; list sample, government; household sample, private sector; and household sample, government. The description includes preparation of the data, selection of the donors, and use of donor and other information to create the item for the recipient.

* * *

The Medical Expenditure Panel Survey (MEPS)

Background

The Medical Expenditure Panel Survey (MEPS) 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. MEPS is cosponsored by the Agency for Healthcare Research and Quality (AHRQ), formerly the Agency for Health Care Policy and Research, and the National Center for Health Statistics (NCHS).

MEPS comprises three component surveys: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC). The HC is the core survey, and it forms the basis for the MPC sample and part of the IC sample. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

MEPS is the third in a series of national probability surveys conducted by AHRQ on the financing and use of medical care in the United States. The National Medical Care Expenditure Survey (NMCES) was conducted in 1977, the National Medical Expenditure Survey (NMES) in 1987. Beginning in 1996, MEPS continues this series with design enhancements and efficiencies that provide a more current data resource to capture the changing dynamics of the health care delivery and insurance system.

The design efficiencies incorporated into MEPS are in accordance with the Department of Health and Human Services (DHHS) Survey Integration Plan of June 1995, which focused on consolidating DHHS surveys, achieving cost efficiencies, reducing respondent burden, and enhancing analytical capacities. To accommodate these goals, new MEPS design features include linkage with the National Health Interview Survey (NHIS), from which the sample for the MEPS-HC is drawn, and enhanced longitudinal data collection for core survey components. The MEPS-HC augments NHIS by selecting a sample of NHIS respondents, collecting additional data on their health care expenditures, and linking these data with additional information collected from the respondents' medical providers, employers, and insurance providers.

Household Component

The MEPS-HC, a nationally representative survey of the U.S. civilian noninstitution-alized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment.

The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a two and a half year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when

combined with other ongoing panels, will provide continuous and current estimates of health care expenditures.

The sampling frame for the MEPS-HC is drawn from respondents to NHIS, conducted by NCHS. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized population, with oversampling of Hispanics and blacks.

Medical Provider Component

The MEPS-MPC supplements and validates information on medical care events reported in the MEPS-HC by contacting medical providers and pharmacies identified by household respondents. The MPC sample includes all hospitals, hospital physicians, home health agencies, and pharmacies reported in the HC. Also included in the MPC are all office-based physicians:

- Providing care for HC respondents receiving Medicaid.
- Associated with a 75 percent sample of households receiving care through an HMO (health maintenance organization) or managed care plan.
- Associated with a 25 percent sample of the remaining households. Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents, including:
- Diagnoses coded according to ICD-9 (9th Revision, International Classification of Diseases) and DSMIV (Fourth Edition, Diagnostic and Statistical Manual of Mental Disorders).
- Physician procedure codes classified by CPT-4 (Current Procedural Terminology, Version 4).
- Inpatient stay codes classified by DRG (diagnosis related group).
- Prescriptions coded by national drug code (NDC), medication names, strength, and quantity dispensed.
- Charges, payments, and the reasons for any difference between charges and payments.

The MPC is conducted through telephone interviews and mailed survey materials.

Insurance Component

The MEPS-IC collects data on health insurance plans obtained through private and public sector employers. Data obtained in the IC include the number and types of private insurance plans offered, benefits associated with these plans, premiums, contributions by employers and employees, and employer characteristics.

Establishments participating in the MEPS-IC are selected through three sampling frames:

- A list of employers or other insurance providers identified by MEPS-HC respondents who report having private health insurance at the Round 1 interview.
- A Bureau of the Census list frame of private-sector business establishments.
- The Census of Governments from the Bureau of the Census.

To provide an integrated picture of health insurance, data collected from the first sampling frame (employers and other insurance providers) are linked back to data provided by the MEPS-HC respondents. Data from the other three sampling frames are collected to provide annual national and State estimates of the supply of private health

insurance available to American workers and to evaluate policy issues pertaining to health insurance. Since 2000, the Bureau of Economic Analysis has used national estimates of employer contributions to group health insurance from the MEPS-IC in the computation of Gross Domestic Product (GDP).

The MEPS-IC is an annual panel survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone follow-up for nonrespondents.

Survey Management

MEPS data are collected under the authority of the Public Health Service Act. They are edited and published in accordance with the confidentiality provisions of this act and the Privacy Act. NCHS provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports and microdata files. Summary reports are released as printed documents and electronic files. Microdata files are released on CD-ROM and/or as electronic files.

Printed documents and CD-ROMs are available through the AHRQ Publications Clearinghouse. Write or call:

AHRQ Publications Clearinghouse
Attn: (publication number)
P.O. Box 8547 Silver Spring, MD 20907
800-358-9295
703-437-2078 (callers outside the United States only)
888-586-6340 (toll-free TDD service; hearing impaired only)

To order online, send an e-mail to: ahrqpubs@ahrq.gov.

Be sure to specify the AHRQ number of the document or CD-ROM you are requesting. Selected electronic files are available through the Internet on the MEPS Web site: http://www.meps.ahrq.gov/

For more information, visit the MEPS Web site or e-mail mepspd@ahrq.gov.

MEPS

Table of Contents

Additional Imputations of Employer Information for the Insurance Component of the Medical Expenditure Panel Survey since 1996

John P. Sommers, PhD, Agency for Healthcare Research and Quality

Background

This report is the second report containing information on imputation of variables for the Insurance Component of the Medical Expenditure Panel Survey (MEPS-IC). The MEPS-IC is a survey of employers, both private industry and public, that collects information on employer-sponsored health insurance. The survey is sponsored by the Agency for Healthcare Research and Quality and conducted by the U.S. Census Bureau. It is designed to collect information on employment-related health insurance, such as premiums and types of plans offered. Information that describes characteristics of the employer is also collected. These data are used to classify employers for calculations of averages and totals and to serve as independent variables for economic modeling.

The sample design of the MEPS-IC is described in Sommers, 1999. Imputation of missing data for 1996 is described in Sommers, 2000a. These documents reflect the survey as of the 1996 survey year. Since that time, unions and insurers of respondents to the Household Component of MEPS (MEPS-HC) have been dropped from the sample due to the low response rates. The sample of self-employed individuals with no employees (SENEs) has also been dropped due to a combination of factors. The major reasons were low response rates and the fact that many self-employed did not have insurance as self-employed individuals, instead they obtained insurance through another employer or through their spouse's employment. Because the employers providing the insurance in these cases are covered through the main sample of employers, this further limited the number of sample persons with usable data. Combined with the low response rates, this caused the sample of SENEs to be of marginal value and so it was dropped.

Since that first survey for 1996, the list of variables imputed with the data have been expanded significantly. Some of these are variables that were not originally collected; however, most are additions to the list that were collected for the 1996 survey year but not imputed. (Copies of the standard 1999 MEPS-IC data collection questionnaires for establishments and plans are available on the MEPS Web site at http://www.meps.ahrq.gov/mepsweb/survey_comp/survey_ic.jsp.)

Because the imputations being described are an expansion of the previous list and the basic technical methods are very similar, this paper will not give the level of detail provided in Sommers, 2000a.

This paper also assumes that the previous imputations described have been completed before the imputations described here take place. For these reasons, readers are advised to be familiar with the previous work described in Sommers, 2000a, in order to have the clearest picture of the process.

General Technical Methods

The original imputation methods report described the process for seven groups of variables. Variables were grouped based on natural relationships. For instance, questions relating to whether an employer offered health insurance to retirees and whether they offered insurance to retirees below age 65, above, or both were done together. Likewise, the variables were ordered to maintain consistency. For instance, type of plan providers was imputed before premiums because premiums are influenced by type of plan providers (Sommers, 2000a).

The basic methods used to produce the additional new imputed results were similar to those used in developing the processes for the first seven sets of imputation. They are grouped and continue in order with later groups built using previous imputations, if required. The new groups are numbered from 8 to 17. As before, each group generally goes through three phases to produce imputed results:

- Data preparation. Data editing is completed and data is normalized; for example, all premium values are annualized.

- Donors are selected for each recipient needing a value. Generally, this is done using a hot-deck method that is similar for all groups. Specifics behind the hot-deck process can be found in Kalton and Kasprzyk, 1986, and are based upon the premise that the expected values of two items are the same if both items agree on a set of important predictive characteristics. The method used to implement this technique was developed by Stiller and Dazell, 1997. It depends on sorting the donors and recipients. As part of the process, class and sort variables are developed and ordered. A donor and recipient must have the same class variables. If donors and recipients disagreed on any class variable then their expected difference in values have been determined to be too large. They can differ on the sort variables, which have less effect on expected values than class variables. Efforts are made to match on sort variables also, but these are ordered so that if not all variables are matched the least important are dropped first in the matching process. (A more extensive description of this process is given in Sommers, 2000a.)

- Final required values are produced. Many times there is a direct substitution of one or more donor values into the recipient's slots for the same variables. However, some of the recipient values are determined by using ratios or other values derived from the donor and applying them to a current recipient value. This is done to maintain consistency among the recipient results. For instance, to obtain the number of employees eligible for health insurance, the donor ratio of eligible to total employees is applied to the number of employees of the recipient. This maintains data relationships. If a direct substitution of the donor's eligible employees were used, the process would need to be limited to donors with very similar values of total employees to the value of the recipient. Otherwise, the recipient values would likely have an expected value that was too high or too low dependent upon the relationship of the employment of the recipient to the average employment of the set of donors.

Process for Each Group of Variables

In the next sections of the report, we proceed through each of the new groups of variables that were imputed for the MEPS-IC. For each group, we give the list of variables to be imputed with reference to their questionnaire name (establishment or plan) and question number. (Copies of the standard 1999 MEPS-IC data collection questionnaires for establishments and plans are available on the MEPS Web site at http://www.meps.ahrq.gov/mepsweb/survey_comp/survey_ic.jsp.) We also describe sort variables used in the selection of all donors for the individual variables within the group, describe class variables used for imputation for all individual variables within the group, and describe the step-by-step process used to create values for each type of recipient from the donor information. Sort and class variables are given for imputation of private sector data. Changes made for imputation of government data are given in the last section of this report. See the appendix for precise definitions of required class and sort variables.

We assume that all logical edits have been performed before the imputation takes place. Thus, for instance, if a respondent gave a total number of part-time employees in question D1b as zero and did not fill in how many were eligible or enrolled, these values would automatically be set to zero. Because of this assumption, we do not discuss logical edits in process descriptions unless this information adds to the discussion of the process.

Throughout the process, we assume a standard definition of a responding establishment and responding plan. An establishment was considered a respondent if it answered that it did or did not provide insurance for its employees, and if the establishment did provide insurance for some of its employees, the establishment also responded at the plan level for at least one of its plans. Responding plans are defined as those that had information provided for at least one of the following items on the plan questionnaire for the specific plan:

- Type of providers, question 2
- Gatekeeper required, question 3
- Purchased or self-insured, question 4
- Plan active enrollment, question 7a
- Premium levels and contributions, questions 8 and 9

In the following sections, we describe the imputations of variables in the new groups 8 through 17. As was done in the previous methods report (Sommers, 2000a), we give (1) the variables imputed in the groups, (2) the sort variables, in order of importance from most to least important, and (3) the class variables. Along with these lists, we give the processes used to convert data from the donor to create the recipient values.

Eighth group of variables

Plan questionnaire (Pq) 6c	Annual plan cost for self-insured plans.
Establishment questionnaire (Eq) E1	Total annual cost of coverage for all hospitalization/physician plans at the location.

Sort variables
None, hot deck process not used.

Class Variables
None

Process
To impute these values does not require the selection of any other donors. At this time, all enrollments and premiums for each plan have been imputed in earlier imputation groups. Using the assumption that the plan premium and enrollments are the same for the entire year at the establishment, then the total annual cost for a plan is the number of single enrollees multiplied by the annual single premium plus the number of married enrollees multiplied by the annual married premium. Using this method directly gives an estimate of total annual plan cost for a self-insured plan. The weighted sum of these estimates by plan for the set of plans collected for the establishment gives an estimate of the total annual cost at the location for hospitalization/physician plans. The weight used is the conditional plan weight within the establishment given the establishment is in the survey.

Ninth group of variables

Eq E8a	Retirees in the firm covered by insurance
Eq E8b	Retirees in the firm with single coverage
Eq E9a	Retiree single coverage premium
Eq E9b	Retiree single coverage employer contribution
Eq E10a	Retiree family coverage premium
Eq E10b	Retiree family coverage employer contribution

(Note these questions do not apply to the establishment in the sample. They apply to the firm that controls the establishment. This is done because retiree data are not available at the establishment level and actually cannot always be related to a specific operating establishment. For instance, retirees within a firm that worked at a closed factory cannot be associated with a particular operating establishment. Thus, retiree questions are for the firm and require special estimation processes to be used in making estimates. For more information, see Sommers, 2000b.)

Sort variables
A different donor is selected for each recipient. A different set of sort variables is used for the hot deck for each variable imputed. This reflects the differing sets of predictors for each value. The variables listed have been placed in the same group because they are all questions concerning retiree coverage, and imputation of some of these variables requires use of one of the previously imputed variables.

For retirees covered by insurance and retirees with single coverage, the sort variables are (1) whether the establishment offers health insurance to retirees under 65, (2) whether the firm offers health insurance to retirees over 65, (3) the industry division group, (4) the industry division, (5) Firm Age Group 2, and (6) Firm Size Class 2. (See the appendix for variable definitions.)

For retiree single coverage premium the sort variables are (1) industry division group, (2) industry division, (3) Firm Size Class 2, (4) Firm Size Class 1, (5) Census division, (6) State, and (7) the size of the retiree family coverage premium.

For retiree single coverage employer contribution the sort variables are (1) industry division group,(2) industry division,(3) Firm Size Class 2, (4) Firm Size Class 1, (5) Census division, and (6) State.

For retiree married coverage premium the sort variables are (1) industry division group, (2) industry division, (3) Firm Size Class 2, (4) Firm Size Class 1, (5) Census division, (6) State, and (7) the size of the retiree single coverage premium.

For retiree married coverage employer contribution the sort variables are (1) industry division group, (2) industry division, (3) Firm Size Class 2, (4) Firm Size Class 1, (5) Census division, and (6) State.

Class variables
The class variables for all the hot deck imputations in this group are Firm Size Class 3 and whether the establishment offered health insurance to retirees.

Process
The values of the retiree single coverage and retiree married coverage premiums are taken directly from the donor establishment. The other four variables are obtained by multiplying a ratio calculated from the donor times a value take from the recipient. The total number of retirees for the firm is the number of employees for the firm of the recipient multiplied by the ratio of the number of retirees from the donor establishment's firm over the total employment of the donor establishment's firm.

The total single enrollees is the total enrollment for the recipient multiplied by the ratio of the total single enrollment for the donor' firm over the total enrollment for the donor's firm.

Each of the two contributions is calculated by multiplying the corresponding (family or single) premium for the recipient by the ratio of the donor plan's corresponding employer contribution over the donor plan's corresponding premium.

Tenth group of variables

Eq E2a   Optional coverages offered
Eq E2b   Total cost of optional coverage

Sort variables
The two variables are imputed in sequence, but the second value does not use a hot-deck routine and no sort values are used. For the first, variable, "optional coverages offered," the file is sorted by industry division and Firm Size Class 1.

Class variables
The variable Firm Size Class 2 is used as a class variable for imputation of the first variable.

Process

The first question if not answered is imputed directly from a donor who provided a response. A donor is an establishment that either checked that it did not offer any optional coverage or checked one or more of the optional coverages listed. A recipient is an establishment that either checked no box or checked that it did not offer coverage and then checked a coverage that was offered.

The total cost of optional coverage for an establishment for those establishments that offered this coverage, whether actual or imputed, and did not report a cost had its costs imputed by applying a factor to its total number of employees enrolled in health insurance.

The factors are derived from the costs of those establishments that reported both the coverages offered and their total costs. Each establishment could offer from one to four coverages. The reporting establishments are grouped by whether they offer one, two, three, or four optional coverages. The weighted sum of the reported optional coverage costs for each group is calculated and divided by the weighted total of their enrolled to obtain a ratio of cost per enrollee for those establishments offering that number of optional coverages.

For each recipient, total costs are determined by multiplying the establishment enrollment by the appropriate factor for the establishment based upon the number of coverages offered by the establishment; i.e., if the establishment offers two optional coverages its total cost is its enrollment times the average reported cost per enrollee for establishments that reported offering two optional coverages.

Eleventh group of variables

Pq 8c   How many former employees are enrolled in plan?

Sort variables
None, hot-deck process not used.

Class variables
Firm Size Class 2 and industry division group

Process
For each cell determined by the two class variables, the weighted sum for reported plans of the number of former employees enrolled was divided by the weighted sum of active enrollees for the reported plans within the same cell. To impute the number of former employees enrolled for a recipient plan, that plan's total enrollment was multiplied by the ratio calculated from reporting donor plans in the same cell.

Twelfth group of variables

Pq 13a	Did plan have a deductible?
Pq 13b	What was annual individual deductible?
Pq 14a	Did the plan require a specific number of individual deductibles be met before the family deductible is met?
Pq 14b	How many family members were required to meet the individual deductible?

Pq 14c	What was the total annual family deducible?
Pq 15a	Was hospital care covered?
Pq 15b	How much and/or what percentage was paid by enrollee for hospital care?
Pq 15c	Is physician care covered?
Pq 15d	How much or what percentage was paid by enrollee for physician care?

Sort variables
The imputation is done in sequence with four hot-deck steps. The variables are imputed in the following order:

- Did the plan have a deductible?
- Was the family deductible a multiple of the single deductible?
- Did the plan have hospital coverage and did the plan have physician coverage?
- The remaining variables in the list.

For the first two hot-deck runs, the files are sorted by (1) Firm Size Class 2, (2) state, and (3) size of the single premium. For the remaining two runs the files are sorted by (1) Firm Size Class 2, (2) type of provider, (3) State, and (3) size of plan single premium.

Class variables
The class variable for whether the plan had a deductible is the type of provider. There are no class variables for the second and third hot-deck runs. The class variables for the fourth hot-deck run are did the plan have a deductible and was the family deductible a multiple of the single deductible.

Process
The variables are related because each set relies on information from the previous imputation either to determine if there needs to be an imputation at that point or to determine a class variable for the next imputation. For instance, we must know if there is a deductible from the first imputation in order to know if there needs to be an imputation for the family deductible. The variables from the first three imputations are needed to determine the structure of the imputation results for the large number of variables imputed in the fourth hot deck. The process approach is given below.

Due to the close interaction of the variables in this group, an important first step is taken using a large number of logical edits. For instance, if a plan has deductibles reported, then it is assumed that the plan had deductibles. If co-pays are given for physician visits, then it is assumed that physician care was covered. Once these edits have been carried out, then the imputation steps are done in a sequence that builds the information in a logical, correlated manner.

The first step is to determine if there was a deductible. A recipient in this group would not have information about whether the plan had a deductible and likely would also not have any information about most of the other variables in this overall group. He/she would certainly not have any information on the type of family deducible imputed in the second hot-deck process nor the levels of the various deductibles in the fourth hot-deck group above.

In this first step, a donor is a plan that had reported whether there was a plan deductible. The value of the donor is directly imputed into the recipient value to determine if the recipient had a deductible.

The donors for the second set are those plans that have a deductible and information on the structure of the family deducible. Recipients lack information on the nature of the family deductible but were known to have had a deductible. As with the first hot-deck step, the donor value is directly imputed into the recipient value.

The third hot deck determines if plans offered physician coverage and/or hospitalization coverage. These two values are determined at this stage of the imputation process because most of the remaining variables in the imputation group are related to one of these types of care. For instance, hospital co-pays must be determined, but before one can determine if there is a hospital co-pay, one must know if there is hospital coverage. If there is no coverage, there is no co-pay.

For this third hot-deck run, the donor plans are all those plans that answered both the questions about type of coverage offered. Recipient plans are those plans that failed to report if the plan covered either hospital coverage, physician coverage or both. The donor value is directly imputed into the recipient value if the recipient needs such a value. For instance, if the recipient plan was reported to have hospitalization coverage, but failed to report about physician coverage, then only the donor's value about physician coverage would be used.

For the fourth hot deck, donors are plans that have all the information for all the items in the group that would be required (for instance, if the plan had no deductible its deductibles could be blank) and that offered family coverage, physician coverage, and hospitalization coverage. Recipients are plans for which it is known whether they have a deductible, which types of coverage are included, and whether a family deductible is a multiple of the single deductible. However, necessary details in these areas are not known. For instance, if the plan had a deductible, not all required deductible values are known. If the plan had hospitalization coverage, then it is not known what the co-pays/percentage paid by the enrollee were.

There is only one donor per recipient. Donors have information for all the possible fields to be imputed. This means a donor is sometimes required to have more than the minimum information required to choose a donor or to provide values for the recipient. For instance, all donors have values for physician co-pays, but the recipient may not require a physician co-pay because the plan does not have physician coverage or the recipient plan has had this co-pay reported. Likewise, the recipient plan may not offer family coverage and thus not require a family deductible, but the donor plan, if it has a deductible, would have family deductibles in case they were needed for the imputation. This completeness and use of the two variables, whether the plan has a deductible and the type of married deductible, assure that the donor plan will have all the information required for any recipient plan in the class. This donor specification was used because (1) almost all reporting plans had the two types of coverage and married coverage, and (2) if any information was given, complete information was given. Thus, (1) very few donors are removed from the imputation by the restriction leaving a large supply of donors, (2) it allows the imputation to be carried out using a simpler process with fewer steps by selecting a single donor for all these variables and then using only the needed information, (3) it helps maintain correlation and consistency of data by using the same

donor, and (4) matches of donor and recipient are still made using the most important prediction variables.

What information is used from the donor and how it is used to provide information for the recipient depends upon the pattern of reported information the recipient plan has. The process is such that a determination is made as to which sections of the recipient plan are missing information, then the process considers each section and the pattern of missing information within that section.

The process handles the remaining items in three parts, all of the deductibles are processed together as a group, the hospital co-pay/percent paid and the physician co-pay/percent paid are each handled a separate groups independent of the other and the deductibles.

The process for the deductibles requires that donor relationships be maintained when imputing values to the recipient using donor ratios of family to single deductibles. How each item is calculated depends upon what values have been reported for the donor and recipient. One must also remember that, at this point in the process, one knows whether the family deductible is a value or a multiple of the single deductible for both the donor and the recipient. Since these variables are class variables in this imputation, the donor and recipient share this characteristic. One also knows whether both the donor and recipient have a deductible. Again, this is because this is a class variable in the process. One also has, if necessary, imputed what type of coverages the recipient provides so one knows if one requires a married deductible or hospitalization deductible or physician deductible. On the other hand, the donor always has each of these three types of coverage and thus can provide for all three types of coverage within its class even if the recipient does not need all three.

Data from donors are used in a way to both retain relationships of the data within the donor and at the same time retain any recipient information available. What is done depends upon the case and what the donor and recipient deductible information is. Some of the key cases, are as follows:

- If the recipient has no deductible, then the deductibles are left empty.

- If the recipient has a deductible and is missing all deductible values, the donor values are simply imputed to the recipient.

- If the recipient requires a family deductible but has a single deductible, one gets the family deductible by multiplying the reported recipient single deductible by the ratio of family to total single deductible of the donor. If the recipient has a family but no single deductible, the process is reversed and the recipient family deductible is divided by the donor ratio to obtain the single deductible.

- To preserve whether there is a single individual deductible or separate deductibles for hospital and physician care, if a single deductible is calculated for the recipient using a donor ratio of total single deductible, and the donor has separate deductibles, then the total recipient value is prorated into separate deductibles using donor values. For instance, if the recipient reported no single deductible but a family deducible of 200, and the donor had a family deductible of 300 and two separate individual deductions of 75, then the total single deductible for the recipient would be 100 =

200*(75 + 75)/300. The 100 would then be prorated to 50 and 50 using the portions 75 and 75 from the donor to allocate the 100 between the two individual deductibles.

- ▪ For cases where the donor and recipient both have family deductibles that are multiples of their single deductions, if the recipient does not have a single deductible, family deducible nor the number of single deductibles required for the family deductible, then all donor values are imputed to the recipient. However, if the recipient has no single deductible but has a number of times the single deductible required for family coverage, then only the recipient single deductible is taken from the donor. As above, if the single deductible of the donor is broken into two separate deductibles, then this pattern would be imputed to the recipient.

- ▪ Only required values are imputed. For instance, if a recipient plan has no family coverage, then no family deductibles are taken from the donor. Thus, if a recipient was missing all values of deductibles but from earlier work one knew that no family coverage was offered, then the family deductible from the donor would not be used for this recipient.

The imputation of co-pays/percent paid are basically direct transfer of values from the donor. Donors for the hospitalization co-pays/percents had a reported value for either the co-pay or the percent. It was assumed that if one were reported and the other value missing that the other value was zero. It was also assumed that if a plan had reported an amount paid that it was per stay unless the donor reported otherwise. Recipients for hospitalization co-pays/percents had both the hospital values missing, but offered hospital coverage. The recipient takes from the donor plan the values of all three variables in the set.

For physician co-pays/percents, the same assumptions and edits were made as for hospitalization values. Thus, for physician co-pay/percents, a recipient plan was a plan with physician coverage and no co-pay or percentage reported, a donor plan had at least one of the two values reported.. As with hospitalization, the recipient values were taken directly from the donor.

Thirteenth group of variables

Pq 17a   Did the plan have a maximum out-of-pocket for an individual and , if so how much?

Sort variables
State

Class variables
Type of provider, Pq 2

Process
The two variables in question Pq 17a are related. Only one of the two should be answered. Donors are those plans with valid responses to the question; that is, either they had no maximum or there was a maximum given. Recipients are those plans that had neither of the two questions answered or both.

Imputation from donor to recipient is by direct substitution of donor to recipient value.

Fourteenth group of variables

Pq 17b   Did the plan have a maximum out-of-pocket for a family and if so how much?

Sort variables
State

Class variables
Type of provider, Pq 2

Process
The two variables are related. Only one of the two should be answered. Donors are those plans with valid responses to the question; that is, either they had no maximum or there was a maximum given. Recipients are those plans that had neither of the two questions or both.

Imputation from donor to recipient is by direct substitution of donor to recipient value.

Fifteenth group of variables

Pq 21   Does plan offer routine outpatient prescription coverage, dental care, orthodontic care (only these three types of coverage are imputed)?

Sort variables
State

Class variables
Type of provider, Pq 2

Process
Donors are those plans that have answered either yes or no to all three of these items. Recipients are those plans without a yes or no answer to all three items. To impute, there is a direct transfer of the donor value for any or all of the three items which are not reported on the recipient plans.

Sixteenth group of variables

Eq E3a   Did the establishment require a waiting period before a new employee could be covered by health insurance?
Eq E3b   If a waiting period was required for 1999, how long was the typical waiting period?

Sort variables
Firm age group, industry division group, SIC2, Firm Size Class 1, State, and establishment size class.

Class variables
None

Process
The process requires two hot-deck imputations. Only establishments that offer health insurance are considered. The donor set for the first hot deck is all establishments that reported whether or not they had a waiting period for health insurance. The recipient set is all establishments that failed to answer whether they required a waiting period for health insurance. Imputation of the recipient value is by direct substitution of the donor value.

The donor set for the second imputation is the set of all establishments that require a waiting period for health insurance for their employees and reported the length of that period. Recipients are those establishments that either reported or had imputed that they had a waiting period for their employees before health insurance coverage began, but failed to report the length of that waiting period. The recipient value of the waiting period length is set equal to the value of the donor's waiting period.

Seventeenth group of variables

Pq 18a	Could the plan have refused to cover persons with certain pre-existing conditions?
Pq 18b	Did this happen in 1999? Pq 19 Did the plan have a policy requiring a waiting period before covering a pre-existing condition?

Sort variables
Census division, state, Firm Size Class 2, and establishment employment

Class variables
Type of provider, Pq 3

Process
The process is completed in three steps. First, all plans that were missing an answer about whether a person could have been refused coverage for a pre-existing condition are recipients. Donors have a reported value for the question. Imputation of a value is direct placement of the donor value into the recipient value.

Plans that could deny coverage due to pre-existing conditions, but did not indicate if this had happened, were recipients for the second value. Donor plans reported if someone had been denied coverage for a condition. Again, imputation was to directly copy the donor value to the recipient.

The third imputation was similar to the first two. Donor plans had a reported value for the question, recipient plans did not. Imputation was by direct transfer of the donor value to the recipient.

Government Imputation Process

The imputation process for sampled governments imputes the same data items as the private sector establishments. The process is similar to that of the private sector. The only differences are sort and class variables used. For government case imputation, the same sort variables are used for all data groups. These are, in sort order, region, state, and government employment size. For government cases, there are no class variables that describe the government. Class variables used are only the specialized variables for plans that apply to that particular imputation group. Thus, for instance, for Group 11, for the private sector, Firm Size Class 2 and industry division class are class variables. For government case imputation, both would be dropped. For the various imputations in Group 12, the variables type of provider, did the plan have a deductible, and was the family deductible a multiple of the single deductible are class variables and are kept for government imputation. The size and industry division variables in Group 11 are dropped because they would describe the government and this type variable is not used.

For the Group 12 imputation, the three variables are kept because they are characteristics of the plan and these variables were used.

References

Kalton, G and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology,* Statistics Canada, Vol. 12, No 1, pp 1–16.

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. 99-037.

Sommers, JP. *Imputation of employer information for the 1996 Medical Expenditure Panel Survey Insurance Component.* Rockville, MD: Agency for Healthcare Research and Quality: 2000. MEPS Methodology Report No. 10. AHRQ Pub. No. 00-0039.

Sommers, JP. "Methods to Produce Establishment and Firm Level Estimates for an Economic Survey." Presented at the International Conference on Establishment Surveys, June, 2000, Buffalo, NY.

Stiller, J and Dalzell, D. (1997) Hot-deck imputation with SAS arrays and macros for large surveys. *Proceedings of the 10th Annual NESUG Conference, North East SAS Users Group,* pp 709–714.

Appendix: Definitions of Selected Variables

Firm Size Class 1
1 if enterprise employment = 0-5
2 if enterprise employment = 6-24
3 if enterprise employment = 25-99
4 if enterprise employment = 100-999
5 if enterprise employment = 1000 or more

Firm Size Class 2
1 if enterprise employment = 0-249
2 if enterprise employment = 250 or more

Firm Size Class 3
1 if enterprise employment = 0-4999
2 if enterprise employment = 5000 or more

Establishment Size Class
1 if establishment employment = 0-10
2 if establishment employment = 11 or more

Industry Division
agriculture if two-digit SIC = 01-09
construction if two-digit SIC = 15-17
retail trade if two-digit SIC = 52-59
mining if two-digit SIC = 10-14
finance, insurance, and real estate if two-digit SIC = 60-67
wholesale trade if two-digit SIC = 50-51
manufacturing if two-digit SIC = 20-39
transportation, communication, and utilities if two-digit SIC = 40-49
services if two-digit SIC = 70-89

Industry Division Group
1 if industry division = agriculture, construction or retail trade
2 if industry division = manufacturing, transportation, communication, utilities or services
3 if industry division = mining, finance, insurance, real estate or wholesale trade

SIC2
The first two digits of the establishment's six digit Standard Industrial Classification (SIC) number.

Firm Age Group
1 if age = 0-16
2 if age = 17 years or more

Firm Age Group 2
1 if age = 0-4
2 if age = 5-9
3 if age = 10-14

4 if age = 15-19
5 if age = 20 years or more

Census Division
New England if State = ME, NH, VT, MA, CT, RI
Mid-Atlantic if State = NY, NJ, PA
East North Central if State = OH, IN, IL, MI, WI
West North Central if State = MN, IA, MO, ND, SD, NE, KS
South Atlantic if State = DE, MD, DC, VA, WV, NC, SC, GA, FL
East South Central if State = KY, TN, AL, MS
West South Central if State = AR, LA, OK, TX
Mountain if State = MT, ID, WY, CO, NM, AZ, UT, NV
Pacific if State = WA, OR, CA, AK, HI

SIC 2
The first two digits of the six-digit SIC code.