MEPS HC-036:
1996-2005 Pooled Estimation File

November 2007

Center for Financing, Access, and Cost Trends
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
301-427-1406

Table of Contents

A. Data Use Agreement

Individual identifiers have been removed from the micro-data contained in these files. Nevertheless, under sections 308 (d) and 903 (c) of the Public Health Service Act (42 U.S.C. 242m and 42 U.S.C. 299 a-1), data collected by the Agency for Healthcare Research and Quality (AHRQ) and/or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which they were supplied; any effort to determine the identity of any reported cases is prohibited by law.

Therefore in accordance with the above referenced Federal Statute, it is understood that:

1. No one is to use the data in this data set in any way except for statistical reporting and analysis; and

2. If the identity of any person or establishment should be discovered inadvertently, then (a) no use will be made of this knowledge, (b) the Director Office of Management AHRQ will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHRQ, and (d) no one else will be informed of the discovered identity; and

3. No one will attempt to link this data set with individually identifiable records from any data sets other than the Medical Expenditure Panel Survey or the National Health Interview Survey.

By using these data you signify your agreement to comply with the above stated statutorily based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates Title 18 part 1 Chapter 47 Section 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

The Agency for Healthcare Research and Quality requests that users cite AHRQ and the Medical Expenditure Panel Survey as the data source in any publications or research based upon these data.

B. Background

1.0 Household Component

The Medical Expenditure Panel Survey (MEPS) provides nationally representative estimates of health care use, expenditures, sources of payment, and health insurance coverage for the U.S. civilian non-institutionalized population. The MEPS Household Component (HC) also provides estimates of respondents' health status, demographic and socio-economic characteristics, employment, access to care, and satisfaction with health care. Estimates can be produced for individuals, families, and selected population subgroups. The panel design of the survey, which includes 5 Rounds of interviews covering 2 full calendar years, provides data for examining person level changes in selected variables such as expenditures, health insurance coverage, and health status. Using computer assisted personal interviewing (CAPI) technology, information about each household member is collected, and the survey builds on this information from interview to interview. All data for a sampled household are reported by a single household respondent.

The MEPS-HC was initiated in 1996. Each year a new panel of sample households is selected. Because the data collected are comparable to those from earlier medical expenditure surveys conducted in 1977 and 1987, it is possible to analyze long-term trends. Each annual MEPS-HC sample size is about 15,000 households. Data can be analyzed at either the person or event level. Data must be weighted to produce national estimates.

The set of households selected for each panel of the MEPS HC is a subsample of households participating in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics. The NHIS sampling frame provides a nationally representative sample of the U.S. civilian non-institutionalized population and reflects an oversample of blacks and Hispanics. MEPS oversamples additional policy relevant sub-groups such as Asians and low income households. The linkage of the MEPS to the previous year's NHIS provides additional data for longitudinal analytic purposes.

2.0 Medical Provider Component

Upon completion of the household CAPI interview and obtaining permission from the household survey respondents, a sample of medical providers are contacted by telephone to obtain information that household respondents can not accurately provide. This part of the MEPS is called the Medical Provider Component (MPC) and information is collected on dates of visit, diagnosis and procedure codes, charges and payments. The Pharmacy Component (PC), a subcomponent of the MPC, does not collect charges or diagnosis and procedure codes but does collect drug detail information, including National Drug Code (NDC) and medicine name, as well as date filled and sources and amounts of payment. The MPC is not designed to yield national estimates. It is primarily used as an imputation source to supplement/replace household reported expenditure information.

3.0 Survey Management

MEPS HC and MPC data are collected under the authority of the Public Health Service Act. Data are collected under contract with Westat, Inc. Data sets and summary statistics are edited and published in accordance with the confidentiality provisions of the Public Health Service Act and the Privacy Act. The National Center for Health statistics (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, micro data files, and tables via the MEPS web site: www.meps.ahrq.gov. Selected data can be analyzed through MEPSnet, an on-line interactive tool designed to give data users the capability to statistically analyze MEPS data in a menu-driven environment.

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Financing Access and Cost Trends, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850 (301-427-1406).

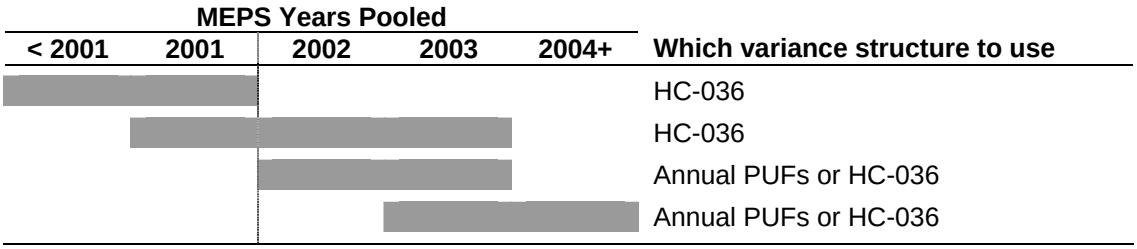
C. Technical and Programming Information

1.0 General Information

To facilitate analysis of subpopulations and/or low prevalence events, it may be desirable to pool together more than one year of MEPS-HC data to yield sample sizes large enough to generate reliable estimates. MEPS-HC samples from year to year are not completely independent because households are drawn from the same sample geographic areas and many persons are in the sample for two consecutive years (see MEPS-HC Methodology Reports for more details at http://www.meps.ahrq.gov). Despite this lack of independence, it is valid to pool multiple years of MEPS-HC data and keep all observations in the analysis because each year of MEPS-HC is designed to be nationally representative. However, to obtain appropriate standard errors when pooling years of MEPS-HC data, it is necessary to specify a common variance structure that properly reflects the complex sample design of the MEPS.

This HC-036 file contains the proper variance structure to use when making estimates from MEPS data that has been pooled over several years. Prior to 2002, each annual MEPS public use file was released with a variance structure unique to the particular MEPS sample in that year. The variance structure in this HC-036 file reconciles the differences in the variance units between the units on the released annual MEPS public use files.

Starting in 2002, the annual MEPS public use files were released with a common variance structure that allows users to pool data from 2002 and forward. This common variance structure starting in 2002 is neither compatible with the variance structure on the annual PUFs released prior to 2002 nor is it compatible with the variance structure on this HC-036 dataset. Therefore, it is only necessary to use the variance structure on this HC-036 dataset when pooling data from MEPS years prior to 2002. The following scenarios provide some guidelines for when analysts should use the variance structure in this HC-036 file.



In the first scenario, only MEPS data from years prior to 2002 are pooled together. In this case, analysts must use the variance structure in HC-036. In the second scenario, data from years prior to 2002 is pooled together with data from 2002 and forward. The variance structure from HC-036 must be used in this circumstance as well. In the last two scenarios, no data from years prior to 2002 are pooled. In both of these cases, analysts may use the variance structure on the released annual public use files (or they may use the variance structure on this HC-036 dataset). In no circumstance should the variance structure on the annual PUFs be combined with the variance structure on the HC-036 dataset.

The variables STRA9605 (stratum of the primary sampling unit) and PSU9605 (primary sampling unit) in this HC-036 dataset provide the appropriate sample design information needed by survey procedures in software packages that implement the with-replacement Taylor series linearization method to obtain estimates of complex sample variances.

The variables BRR1 – BRR64 in the HC-036BRR dataset provide a comparable replicate sample design structure used by survey procedures in software packages that implement the balanced repeated replication (BRR) method to obtain estimates of complex sample variances.

2.0 Data File Information

Released as an ASCII data file (with SAS® and SPSS® user statements) and in SAS Transport version, the HC-036 file contains 169,717 records corresponding to the number of unique persons in MEPS Panels 1 through the first year of Panel 10. These records contain the standard MEPS-HC person-level ID variables (DUPERSID + PANEL), as well as the pooled variance estimation structure (STRA9605 and PSU9605).

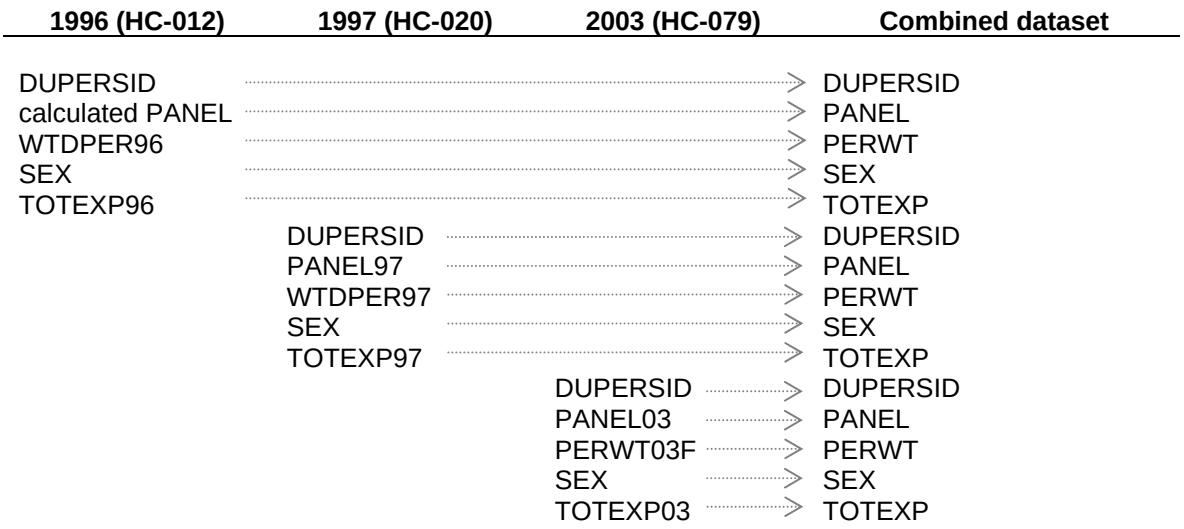
There is a record for each unique person appearing in any of the 1996-2005 MEPS HC full-year person level public use files: HC-012, HC-020, HC-028, HC-038, HC-050, HC-060, HC-070, HC-079, HC-089 and HC-097. These ten data sets have a combined total of 306,238 records; however, as each person may appear in one or two of these data sets, the number of unique persons (169,717) is fewer than the number of records on the annual files (306,238).

3.0 Linking Instructions

The following steps should be taken to create a pooled analysis dataset over any combination of years of the MEPS.

1. Create a dataset for each year containing the person- and/or event-level records of all persons to be included in the analysis. Keep the unique person identifier (DUPERSID and PANEL), the person-level sampling weight, any classification variables (e.g., sex, race/ethnicity) and response variables (e.g., total expenditure amount, number of prescription drug purchases, etc) to be used in the data analysis.

2. Reconcile the discrepancies in variable names. For all years, most variable names on the annual public use files contain a 2-digit year suffix. For instance, in the 1997 consolidated person-level file (HC-020) the panel variable is called PANEL97, the total annual expenditure amount variable is called TOTEXP97 and the sampling weight variable is called WTDPER97. But in the 2003 dataset (HC-079) these same variables are named PANEL03, TOTEXP03 and PERWT03F, respectively, and in the 1996 dataset (HC-012) the total expenditure and sampling weight variables are named TOTEXP96 and WTDPER96, respectively, and the panel variable is missing (users should assign a value of 1 for each record in HC-012). As illustrated below, the variable names must be made consistent before pooling the data.



3. Create a pooled analysis dataset by combining the individual-year datasets by row; that is, append the records from the 1996 dataset with those from the 1997 and 2003 datasets.

4. Attach the pooled variance structure to the pooled analysis dataset by column; that is, merge the variables STRA9605 and PSU9605 from this HC-036 to the pooled analysis dataset by DUPERSID and PANEL keeping all records in the pooled analysis dataset and only those records in HC-036 that match. Depending on the software being used to manage the datasets, the pooled analysis dataset may need to be sorted by DUPERSID and PANEL prior to merging.

4.0 Other Considerations

When working with pooled data, analysts should consider whether they need to adjust the survey weights from the annual public use files to account for the reprojection of survey estimates to a multi-year time period. The survey weights provided in the 1996 annual dataset (HC-012) project the HC-012 sample to the US population in 1996, and the survey weights in the 1997 dataset (HC-020) project the HC-020 sample to the US population in 1997. When combining two years of annual MEPS data (e.g., 1996 and 1997), these single-year weights over-represent the population in the new two-year period (1996 and 1997) by a factor of 2. Likewise, when combining three years of MEPS data, the single-year weights over-represent the new three-year population by a factor of 3.

This over-representation will only the affect estimates of totals but not the estimates of proportions. That is, all estimates of total expenditures and their standard errors will be twice as high as they should be if using the annual weights on the annual public use files pooled over two years without adjustment; these same estimates will be three times too high if pooling over three years. Ratio estimates, such as the mean expenditure or the percent of expenditures paid out of pocket, will not be too high when using the annual weights after pooling several years of MEPS datasets together. Users wishing to estimate totals have two options to account for the multi-year period: they may factor the weights before they make any estimates or they may factor the estimates themselves (they should not do both).

- To illustrate the first method (factoring the weights), users who pool two years of MEPS data should divide the sampling weight (variable PERWT if following the example above) by 2 prior to making any estimates. They would divide the sampling weight by three if pooling three years of MEPS data together. With this adjustment to the sampling weight, all estimates of totals (and their standard errors) will reflect the new multi-year period. Estimates of proportions (and their standard errors) will also be correct after this adjustment.
- To illustrate the second method (factoring the estimates themselves), users would make the estimates of totals (and optionally their proportions) with the annual weights as is. They would then factor the estimates of the totals (as well as their standard errors) by the number of years that were pooled. If the estimates were made with two years of data, the totals and their standard errors would be divided by 2, if they were made with three years of data, the totals and their standard errors would be divided by 3. Users would only adjust the estimates and standard errors of totals, not those of proportions.

When pooling data over several years of the MEPS for the purpose of increasing the sample size for a small subdomains of the population (e.g., obtaining the total and mean expenditures for prescription drugs among children with asthma), users must be careful to maintain the integrity of the MEPS survey design. The MEPS design is accounted for by the full set of survey stratum and PSU values in the HC-036 dataset or by the full set of replicate variables in the HC-036BRR dataset. Survey procedures must have this information in order to properly account for the complex sample design of the MEPS in any estimates they produce. When users subset records from the annual datasets of only those respondents in the subdomain of interest (e.g., children with asthma), it is very unlikely that they will have enough values of stratum and PSU to properly account for the MEPS survey design in any linearized estimate of the sampling variances. Users have two choices to properly account for the MEPS survey design when making estimates for small subdomains.

- The first option is to use the BRR method of variance estimation rather than the with-replacement linearization method. In this method, users may legitimately subset to the subdomain of interest because the MEPS sample design is accounted for by the set of 64 BRR replicate values.
- The second option is to extract records from all respondents, not just those in the subdomain of interest, but then to construct a flag to identify which respondents are in the subdomain. Users would use the with-replacement design option on the survey procedure and specify the subdomain using the flag.

Users should consult the manual of the software package they are using as the way in which survey designs and subdomains are specified are specific to each particular software.

5.0 Further Information

For any question regarding the HC-036 file or pooling of data, please contact Fred Rohde by email at: frederick.rohde@ahrq.hhs.gov.