

**MEPS HC-036:  
1996-2003 Pooled Estimation File**

**October 2005**

**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 .....	A-1
B. Background .....	B-1
1.0 Household Component.....	B-1
2.0 Medical Provider Component .....	B-2
3.0 Insurance Component .....	B-2
4.0 Survey Management .....	B-3
C. Technical and Programming Information .....	C-1
1.0 General Information.....	C-1
2.0 Data File Information.....	C-1
3.0 Linking Instructions .....	C-1
4.0 Other Considerations .....	C-2
5.0 Further Information.....	C-3

## A. Data Use Agreement

Individual identifiers have been removed from the microdata contained in the files on this CD-ROM. 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.
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.
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**

This documentation describes one in a series of public use files from the Medical Expenditure Panel Survey (MEPS). The survey provides a new and extensive data set on the use of health services and health care in the United States.

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) and the National Center for Health Statistics (NCHS).

MEPS is a family of three surveys. The Household Component (HC) is the core survey and forms the basis for the Medical Provider Component (MPC) and part of the Insurance Component (IC). 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, and the National Medical Expenditure Survey (NMES) was conducted in 1987. Since 1996, MEPS has continued 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 advance these goals, MEPS includes linkage with the National Health Interview Survey (NHIS)—a survey conducted by NCHS 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.

### **1.0 Household Component**

The MEPS HC, a nationally representative survey of the U.S. civilian noninstitutionalized 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 2½-year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and

use for 2 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. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized population, with oversampling of Hispanics and blacks.

## **2.0 Medical Provider Component**

The MEPS MPC supplements and/or replaces 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 home health agencies and pharmacies reported by HC respondents. Office-based physicians, hospitals, and hospital physicians are also included in the MPC but may be subsampled at various rates, depending on burden and resources, in certain years.

Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents. The MPC is conducted through telephone interviews and record abstraction.

## **3.0 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 identified by MEPS HC respondents) are linked back to data provided by those respondents. Data collected from the two Census Bureau sampling frames are used to produce annual national and State estimates of the supply and cost of private health insurance available to American workers and to evaluate policy issues pertaining to health insurance. National estimates of employer contributions to group health insurance from the MEPS IC are used in the computation of Gross Domestic Product (GDP) by the Bureau of Economic Analysis.

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 followup for nonrespondents.

#### **4.0 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, microdata files, and compendiums of tables. Data are also released through MEPSnet, an online interactive tool developed to give users the ability to statistically analyze MEPS data in real time. Summary reports and compendiums of tables are released as printed documents and electronic files. Microdata files are released on CD-ROM and/or as electronic files.

Printed documents and selected public use file data on 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  
410-381-3150 (callers outside the United States only)  
888-586-6340 (toll-free TDD service; hearing impaired only)

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/>

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 data to yield sample sizes large enough to generate reliable estimates. MEPS 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 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 data and keep all observations in the analysis because each year of MEPS is designed to be nationally representative. However, to obtain appropriate standard errors when pooling years of MEPS data, it is necessary to specify a common sample design structure.

This file contains the variables required to put all data from 1996-2003 into a common structure required to produce appropriate standard error estimates when pooling multiple years of MEPS data. Specifically, the file contains the 2 variables (STRA9603 and PSU9603) that should be used to specify the complex survey design structure when conducting analyses of pooled data using complex survey software packages that apply the Taylor Series approach (e.g., SUDAAN and STATA). It is only necessary to use this file when one or more years of data being pooled are from 1996 through 2001 because MEPS annual files are being produced with a standardized sample design structure for 2002 and beyond. Therefore, when all the years being pooled are for 2002 and later it will be sufficient to pool the individual annual files without linking to this file<sup>1</sup>.

### 2.0 Data File Information

Released as an ASCII data file (with SAS and SPSS user statements) and in SAS Transport version, the H36 file contains 133,885 records. These records contain the standard MEPS person-level ID variables (DUID, PID and DUPERSID), as well as the 1996-2003 combined variance stratum and PSU variables (STRA9603 and PSU9603, respectively).

There is a record for each person who appears on any of the 1996-2003 MEPS full-year person level public use files: HC-012, HC-020, HC-028, HC-038, HC-050, HC-060, HC-070, and HC-079. These seven data sets have a combined total of 237,874 records. However, because there are panels of households which appear in more than one year, there are duplicate persons over the seven years and thus fewer persons than records.

### 3.0 Linking Instructions

The following steps should be taken to create a pooled file over the seven years.

---

<sup>1</sup> In public use files for 2002 and beyond, the survey design variables are named VARSTR AND VARPSU.

1. Create a SAS file for each year containing persons desired from that year of data. Keep the person identifier DUPERSID and any other variables desired for the analysis.
2. Rename similar variables from different years to common names. As part of this process one may have to rename variables from different years so that they have a consistent name. Many variables have names for each year that are of the form nameyy where the name portion is the same, but yy represents the year of the file. To use as the same variable one must rename each year as the same variable. For instance, name96 and name97 might be given the title NAME.
3. Concatenate files together from the seven years.
4. Sort the file of combined years by DUPERSID.
5. Merge the combined file with the HC-036 data file by DUPERSID and keep records that are in both files. Retain all variables.
6. Produce desired estimates as desired variables STRA9603 and PSU9603 in the same manner as the variables VARSTRyy and VARPSUyy are used in variance calculations for individual years.

#### **4.0 Other Considerations**

**In working with pooled data one must consider the analysis goals and tailor weights and combine person records to fit the analysis plan. However, for many types of work one does not need to combine records of the same person (persons with the same value of DUPERSID). For instance, if one is making estimates of totals for 8 years or an average over the 8 years, one can retain the individual person records from the 8 years of data. Likewise, if one were running a linear regression model, there would be no need to combine records by person.**

The fact that some individuals have two records on a pooled data set is generally not a problem. In fact, if one is calculating annual population and subpopulation means and rates, the two records are required in the analysis. This is because each record is part of the sample for a different year and to calculate valid averages and rates over the combined years, all records for each of the individual years must be included on the file. Any correlation in the data set caused by the multiple records is taken into account by the specialized software used to calculate errors for complex survey samples.

The only time one should use only one record for an individual is when the analysis is of a variable that is constructed using both years a person was in the sample. For instance, one record should be used if a change in behavior or health over the two years is being analyzed. (One can use the rule of thumb, that if the variable can only have one value over the two years the person was in the sample and requires information from both years, then there should only be one record per individual.)



Generally, if one is calculating annual averages or rates using multiple years of data, the weight used should be the standard full year weight for that record divided by the number of years of data being pooled.

For analysis of the latter type of longitudinally constructed variables discussed above, one should use the weights for the person from the longitudinal file for the panel to which the person belongs, divided by the number of panels of individuals that have been pooled for the analysis.

## **5.0 Further Information**

For any question regarding the HC-036 file or pooling of data, please contact John Sommers by email at: [jsommers@ahrq.gov](mailto:jsommers@ahrq.gov).