

**MEPS HC-036:
1996-2024 Pooled Linkage
Variance Estimation File**

August 2026

**Agency for Healthcare Research and Quality
Center for Financing, Access, and Cost Trends
5600 Fishers Lane
Rockville, MD 20857**

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-1
3.0 Survey Management	B-2
C. Technical and Programming Information	C-1
1.0 General Information.....	C-1
2.0 Data File Information.....	C-2
3.0 Linking Instructions	C-2
4.0 Adjustment of Analytic Weight Variable	C-4
5.0 Subpopulation Analysis Caveat	C-4
6.0 Further Information.....	C-4

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 dataset 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) Director - Office of Management Services 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 datasets other than the Medical Expenditure Panel Survey or the National Health Interview Survey. Furthermore, linkage of the Medical Expenditure Panel Survey and the National Health Interview Survey may not occur outside the AHRQ Data Center, NCHS Research Data Center (RDC) or the U.S. Census RDC network.

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.

AHRQ 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, payment sources, 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 survey's panel design usually includes 5 Rounds of interviews spanning 2 full calendar years. Note that due to the difficulty of data collection during the COVID-19 pandemic, interview completion rates in 2020 and 2021 decreased abruptly. To balance this and increase the number of completes to be comparable to previous years, Panel 23 was extended into 2020 and 2021 with nine rounds of data collection. Panel 24 was extended into 2022 with nine rounds of data collection. Computer-assisted personal interviewing (CAPI) technology or computer-assisted video interviewing (CAVI) technology is used to collect information about each household member, which the survey builds on from interview to interview. A single household respondent reports all data for a sampled household.

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. Historically, each annual MEPS HC sample consists of up to 15,000 households. Data can be analyzed at the person, the family, or the 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 NCHS. The NHIS sampling frame provides a nationally representative sample of the U.S. civilian noninstitutionalized population. In 2006, the NCHS implemented a new NHIS sample design that included households with Asian persons in addition to households with Black and Hispanic persons in the minority group oversampling. In 2016, NCHS introduced another sample design that discontinued the oversampling of these minority groups.

2.0 Medical Provider Component

When the household CAPI instrument is completed and permission is obtained from the household respondents to contact their medical provider(s), a sample of these providers is contacted by telephone to obtain information that household respondents cannot accurately provide. This part of MEPS is called the Medical Provider Component (MPC) and it collects information on dates of visits, diagnosis and procedure codes, and charges and payments. The Pharmacy Component (PC), a subcomponent of the MPC, does not collect data on charges or on diagnosis and procedure codes, but it does collect detailed information on drugs, including the National Drug Code (NDC) and medicine name, as well as payment amounts. The MPC is not designed to yield national estimates. It is primarily used as an imputation source to supplement or 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. The MEPS HC data are collected under contract with Westat, and the MEPS MPC data are collected under contract with RTI International. Datasets and summary statistics are edited and published in accordance with the confidentiality provisions of the Public Health Service Act and the Privacy Act. The NCHS provides consultation and technical assistance.

As soon as the MEPS data are collected and edited, they are released to the public in stages of microdata files and tables via the [MEPS website](#) and [AHRQ Data Tools site](#).

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, AHRQ, 5600 Fishers Ln, Rockville, MD 20857.

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 (i.e. combine) more than one year of MEPS HC data to yield sample sizes large enough to generate reliable estimates. MEPS HC samples in most years 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 <https://meps.ahrq.gov/mepsweb/>). 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 the 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 have been pooled over multiple years, specifically when one or more years are from 1996-2001 or 2019-2024. Prior to 2002, each annual MEPS public use file was released with a variance structure unique to the particular MEPS sample in that year. Also, the years 2019 and later have a common variance structure but slightly different than earlier years. 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.

Between 2002 and 2018, the annual MEPS public use files were released with a common variance structure that allows users to pool data from 2002 to 2018. Also, for 2019 and later years, the MEPS public use files were released with another common variance structure that allows users to pool data from 2019 and later years. However, the common variance structure for 2002 to 2018 and the common variance structure for 2019 and later years are not compatible with each other nor the structure on the annual PUFs released prior to 2002. Therefore, it is necessary to use the variance structure on this HC-036 dataset when pooling data from MEPS years prior to 2002 and also from 2019 - 2024. The following image presents some scenarios to clarify when analysts should use the variance structure in this HC-036 file.

MEPS Years Pooled						Which variance structure to use
< 2001	2001	2002	2003	2004- 2018	2019- 2024	
[Shaded]						HC-036
	[Shaded]					HC-036
						Annual PUFs
			[Shaded]			Annual PUFs
					[Shaded]	Annual PUFs
			[Shaded]			HC-036

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 are pooled together with data from 2002 and forward up to 2018. The variance structure from HC-036 must be used in this circumstance as well. In the third and fourth scenarios, no data from years prior to 2002 or from 2019 to 2024 are pooled. In both of these cases, analysts should use the variance structure on the released annual public use files. In the fifth scenario, only data from 2019 to 2024 are pooled, so analysts should use the variance structure on the released annual public use files. In the last scenario, data from year 2019-2024 are pooled with other years. The variance structure from HC-036 must be used in this circumstance as well. In no circumstance should the variance structure on the annual PUFs be combined with the variance structure on the HC-036 dataset.

The variables STRA9624 (stratum of the primary sampling unit) and PSU9624 (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 – BRR128 in the [HC-036BRR dataset](#) provide a comparable replicate sample design structure. These replicates can be incorporated in software package survey procedures that implement the balanced repeated replication (BRR) method to produce estimates of complex sample variances.

2.0 Data File Information

Released as an ASCII data file (with SAS®, Stata®, SPSS®, and R programming statements and data user information) and as a SAS dataset, SAS transport file, Stata dataset, and Excel file, the HC-036 file contains 483,346 records corresponding to the number of unique persons in MEPS from 1996-2024, with the exception of Panel 22 persons who appear in both the 2017 and 2018 HC files. The format for DUPERSID changed in 2018 requiring that HC-036 include these persons twice, one time with their 2017 DUPERSID and then again with their 2018 DUPERSID. All records contain the standard MEPS HC person level ID variables (DUPERSID + PANEL), as well as the pooled variance estimation structure (STRA9624 and PSU9624).

There is a record for each unique person appearing in any of the 1996-2024 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, HC-097, HC-105, HC-113, HC-121, HC-129, HC-138, HC-147, HC-155, HC-163, HC-171, HC-181, HC-192, HC-201, HC-209, HC-216, HC-224, HC-233, HC-243, HC-251, and HC-256. These data sets have a combined total of 897,782 records; however, as each person may appear in one or two of these data sets, the number of records with unique DUPERSID in HC-036 (483,346) is fewer than the total number of records on the annual files.

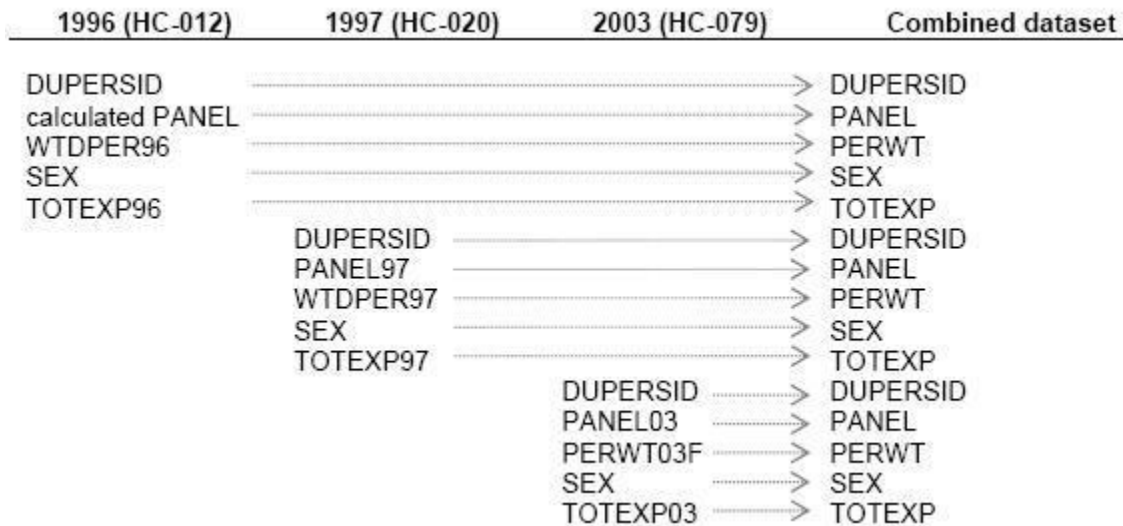
3.0 Linking Instructions

The following steps should be taken to create a pooled analysis dataset.

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). Starting in 2005, the panel variable is simply named PANEL (no year suffix). As illustrated below, the variable names must be made consistent before pooling the data.



3. Create a pooled analysis dataset by simply combining the individual-year datasets (e.g., the records from the 1996 and 1997 files). In other words, the number of records in the pooled file will equal the sum of the record counts for the individual annual files being pooled.
4. Attach the pooled variance structure to the pooled analysis dataset by merging the variables STRA9624 and PSU9624 from this HC-036 file to the pooled analysis dataset by DUPERSID and PANEL keeping all records in the pooled analysis dataset only. 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. This step will add two additional variables to the pooled file (STRA9624 and PSU9624) but have no impact on the number of records.

4.0 Adjustment of Analytic Weight Variable

It is generally recommended that analysts adjust the analytic weight variable by dividing it by the number of years being pooled. The sum of these adjusted weights represents the average annual population size for the pooled period (rather than the sum of the population sizes across multiple years that would result from unadjusted weights). Although this adjustment will have no effect on estimated means, proportions or regression coefficients because the weight variable is being divided by a constant (i.e. number of years), estimates of totals based on adjusted weights will reflect an “average annual” basis rather than the entire pooled period. On the other hand, if the objective is to produce an estimated total for the entire pooled period (e.g. total medical expenditures across multiple years rather than average per year), then the analytic weight variable should not be divided by the number of years in the pooled period.

5.0 Subpopulation Analysis Caveat

When pooling data over several years to increase sample sizes for 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 on both the annual files and this HC-036 pooled linkage variance estimate file.¹ When users create analytic subfiles that contain only respondents in the subdomain of interest (e.g., children with asthma), it is very unlikely that there will be all combinations of stratum and PSU that properly account for the MEPS survey design in a linearized estimate of the sampling variances. Therefore, the following approach is recommended for analyzing subpopulations in MEPS:

- 1) Construct a flag variable for all survey respondents that can be used to identify persons in the subdomain of interest,
- 2) Using a with-replacement design option for a Taylor Series procedure in a complex survey design statistical software package, read in records from all respondents (i.e., not just those in the subdomain of interest) and specify the analytic subdomain using the flag variable (see step 1 above).²

6.0 Further Information

For any question regarding the HC-036 file or pooling of data, please contact Sadeq Chowdhury by email at: sadeq.chowdhury@ahrq.hhs.gov .

1 The MEPS design is also accounted for by the full set of replicates in the [HC-036BRR data set](#)

2 The syntax for specifying survey designs and analytic subdomains varies across software packages (see section IB at https://meps.ahrq.gov/mepsweb/survey_comp/clustering_faqs.jsp for examples).