

**MEPS HC-244**  
**Panel 26 Longitudinal Data Public Use File**  
**September 2024**

**Agency for Healthcare Research and Quality  
Center for Financing, Access, and Cost Trends  
5600 Fishers Ln  
Rockville, MD 20857  
(301) 427-1406**

## **Table of Contents**

<b><u>Section</u></b>	<b><u>Page</u></b>
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 and Data Collection .....	B-2
C. Technical and Programming Information .....	C-1
1.0 General Information.....	C-1
2.0 Data File Information.....	C-1
2.1 Variables.....	C-2
2.1.1 Variables from Annual Full-Year Consolidated Files.....	C-2
2.1.2 Constructed Variables for Selection of Group .....	C-6
2.1.3 Estimation Variables .....	C-7

## 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:

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

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

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.

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 healthcare. Estimates can be produced for individuals, families, and selected population subgroups. The panel design of the survey includes 5 rounds of interviews covering 2 full calendar years. Additional rounds were added to Panel 24 in 2020 and 2021, covering third and fourth years, respectively, to compensate for the smaller number of completed interviews in later panels. These extra rounds provide data for examining person-level changes in selected variables such as expenditures, health insurance coverage, and health status. Information about each household member is collected through computer-assisted personal interviewing (CAPI) technology, 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. Historically, each annual MEPS HC sample consists of approximately up to 15,000 households. Data can be analyzed at the person, the family, 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 (NCHS). The NHIS sampling frame provides a nationally representative sample of the U.S. civilian noninstitutionalized population. In 2006, the NCHS implemented a new sample design for the NHIS, to include households with Asian persons in addition to households with Black and Hispanic persons in the oversampling of minority populations. In 2016, NCHS introduced another sample design that discontinued the oversampling of these minority groups.

### **2.0 Medical Provider Component**

When the household CAPI interview is completed, and permission is obtained from the household survey respondents to contact their medical provider(s), a sample of these providers is contacted by telephone to obtain information that household sample members cannot accurately provide. This part of the 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 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 and Data Collection**

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, Inc. and the MEPS MPC data are collected under contract with Research Triangle Institute. 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 [datatools.ahrq.gov](http://datatools.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, 5600 Fishers Lane, Rockville, MD 20857 (301-427-1406).

## C. Technical and Programming Information

---

### 1.0 General Information

This documentation describes the Panel 26 Longitudinal Public Use File (PUF) from the MEPS HC. It was released as an ASCII file (with related SAS, Stata, SPSS, and R programming statements and data user information) and as a SAS dataset, a SAS transport dataset, a Stata dataset, and an Excel file. The Panel 26 Longitudinal PUF provides information collected from a nationally representative sample of the U.S. civilian noninstitutionalized population for the two-year period 2021-2022. The file contains 2,737 variables and has a logical record length of 7,756 with an additional 2-byte carriage return/line feed at the end of each record.

This file consists of MEPS survey data obtained in Rounds 1-5 of MEPS Panel 26 and can be used to analyze changes over a two-year period. Variables in the file pertaining to survey administration, demographics, employment, health status, disability days, quality of care, patient satisfaction, health insurance and medical care use and expenditures were obtained from the MEPS 2021 and 2022 Full-Year Consolidated PUFs (HC-233 and HC-243, respectively).

The following documentation offers a brief overview of the contents and structure of the files and programming information. A codebook of all the variables included in the Panel 26 Longitudinal PUF is provided in a separate file (H244CB.PDF). A database of all MEPS products released to date and a variable locator indicating the major MEPS data items on public use files that have been released to date can be found on the [MEPS website](#).

### 2.0 Data File Information

The Panel 26 Longitudinal PUF contains records for 6,741 persons in Panel 26 who were respondents for the period they were in-scope for the survey (i.e., a member of the civilian non-institutionalized population) during the two-year period. Only persons with positive person-level weights (PERWT21F or PERWT22F) are included in the longitudinal PUF data. Data are available for all five rounds for 93.38% of the cases (6,295). The remaining 6.62% (446 persons) do not have data for one or more rounds but were in-scope for all rounds they participated in the survey. These persons are those who were born, died, were in the military or an institution, or left the country during the two-year period. In contrast, persons in the panel who participated in the survey for only part of the period they were in scope are not included in this file. To compensate for this attrition, adjustments were made in the construction of the panel weight variable included in this file (LONGWT). The codebook provides both weighted and unweighted frequencies for each variable on the data file. The LONGWT variable should be used to produce national estimates for the two-year period.

## 2.1 Variables

### 2.1.1 Variables from Annual Full-Year Consolidated Files

Most variables on this file were obtained from the MEPS 2021 and 2022 Full-Year Consolidated PUFs (HC-233 and HC-243, respectively). However, names for time dependent variables from these files are modified in order to: 1) eliminate duplicate variable names for data reflecting different time periods during the panel, and 2) standardize variable names to facilitate pooling of multiple MEPS panels for analysis.<sup>1</sup> Generally, annual variables with a suffix of “21” and “22” are renamed with a suffix of “Y1” and “Y2”, respectively. Variables with a suffix of “31”, “42”, and “53” are renamed with a suffix denoting the round the data was collected (i.e., “1”, “2” or “3” for variables originating from Rounds 1-3 on the 2021 full-year file and “3”, “4”, or “5” for variables originating from Rounds 3-5 on the 2022 full-year file).<sup>2</sup> It is necessary to use this crosswalk in conjunction with documentation for the 2021 and 2022 Full-Year Consolidated PUFs to obtain a full description of variables on this file. Table 1 below provides the crosswalk summarizing the scheme used for renaming variables from the annual files.

Table 1. Crosswalk of Variable Names between the Full-Year Consolidated PUFs and the Longitudinal PUF

Type of Variable	Full-Year Consolidated PUF Variable Name Suffix	Longitudinal PUF Variable Name Suffix	Specific cases or examples
Constant (i.e., not round or year specific)	No suffixes	No suffixes	All variables: BORNUSA=BORNUSA DOBMM=DOBMM DOBY=DOBY DATAYEAR=DATAYEAR DUID=DUID PID=PID DUPERSID=DUPERSID EDUCYR=EDUCYR HIDE=HIDE HISPANX=HISPANX HISPNCAT=HISPNCAT HWELLSPK=HWELLSPK INTVLANG=INTVLANG OTHLGSPK=OTHLGSPK PANEL=PANEL PID=PID RACEAX=RACEAX RACEBX=RACEBX

<sup>1</sup> A variable named PANEL is also included to facilitate pooling across panels. This variable is simply the panel number and is therefore constant across all records within a longitudinal file. The ten-character variable DUPERSID uniquely identifies each person represented on the file and is the combination of the variables DUID (PANEL + Dwelling Unit ID) and PID (Person Number).

<sup>2</sup> While Round 3 values were obtained for most observations from the 2022 Full Year Consolidated PUF, they were obtained from the 2021 Full Year Consolidated PUF for sample persons where YEARIND=2 (i.e., in 2021 only).

Type of Variable	Full-Year Consolidated PUF Variable Name Suffix	Longitudinal PUF Variable Name Suffix	Specific cases or examples
			RACEWX=RACEWX RACEV1X=RACEV1X RACEV2X=RACEV2X RACETHX=RACETHX SEX=SEX VARPSU=VARPSU VARSTR=VARSTR WHTLGSPK=WHTLGSPK YRSINUS=YRSINUS
Annual, family related variables	YR	Y1 or YR1  Y2 or YR2	All variables: FAMIDYR=FAMIDYR1 (2021 file) FAMRFPYR=FAMRFPY1 (2021 file) FAMSZEYR=FAMSZYR1 (2021 file)  FAMIDYR=FAMIDYR2 (2022 file) FAMRFPYR=FAMRFPY2 (2022 file) FAMSZEYR=FAMSZYR2 (2022 file)
Annual, CPS family identifiers	No suffix	Y1  Y2	All variables: CPSFAMID= CPSFAMY1 (2021 file)  CPSFAMID= CPSFAMY2 (2022 file)
Annual, health insurance eligibility units	No suffix	Y1  Y2	All variables: HIEUIDX=HIEUIDY1 (2021 file)  HIEUIDX=HIEUIDY2 (2022 file)
Annual, in-scope variables	No suffixes	YR1  YR2	All variables: INSCOPE=INSCPYPY1 (2021 file)  INSCOPE=INSCPYPY2 (2022 file)
12/31 status variables	1231 in 2021 file  1231 in 2022 file	Y1  Y2	All variables: FAMS1231=FAMSY1 (2021 file) FCRP1231=FCRPY1 (2021 file) FCSZ1231= FCSZY1 (2021 file) FMRS1231= FMRSY1 (2021 file) INSC1231=INSCY1 (2021 file)  FAMS1231=FAMSY2 (2022 file) FCRP1231=FCRPY2 (2022 file) FCSZ1231= FCSZY2 (2022 file) FMRS1231= FMRSY2 (2022 file) INSC1231=INSCY2 (2022 file)
Annual	21, 21X, 21F, or 21C  22, 22X, 22F, or 22C	Y1, Y1X, Y1F, or Y1C  Y2, Y2X, Y2F, or Y2C	Examples: TOTEXP21=TOTEXPY1 AGE21X=AGEY1X  TOTEXP22=TOTEXPY2 AGE22X=AGEY2X





Type of Variable	Full-Year Consolidated PUF Variable Name Suffix	Longitudinal PUF Variable Name Suffix	Specific cases or examples
	53_Myy in 2022 file		ADRNK442_M20=ADRNK44_M20 DENTIN53_M23=DENTIN5_M23
Diabetes preventive care	2053, 2153, and 2253 in 2021 file  2153, 2253, and 2353 in 2022 file	Y0R3 for 2020 Y1R3 for 2021 Y2R3 for 2022  Y1R5 for 2021 Y2R5 for 2022 Y3R5 for 2023	Examples: DSEB2053=DSEBY0R3 (2021 file) DSEY2053=DSEYY0R3 (2021 file) DSEY2153=DSEYY1R3 (2021 file) DSEY2253=DSEYY2R3 (2021 file)  DSEB2153=DSEBY1R5 (2022 file) DSEY2153=DSEYY1R5 (2022 file) DSEY2253=DSEYY2R5 (2022 file) DSEY2353=DSEYY3R5 (2022 file)
Job Change	3142 or 4253	12 for 2021 23 for 2021  34 for 2022 45 for 2022	All cases: CHGJ3142=CHGJ12 (2021 file) CHGJ4253=CHGJ23 (2021 file) YCHJ3142=YCHJ12 (2021 file) YCHJ4253=YCHJ23 (2021 file)  CHGJ3142=CHGJ34 (2022 file) CHGJ4253=CHGJ45 (2022 file) YCHJ3142=YCHJ34 (2022 file) YCHJ4253=YCHJ45 (2022 file)
Cancer/ Cancer in remission <sup>4</sup>	No suffixes <sup>5</sup>	Y1 for 2021  Y2 for 2022	Example: CALUNG=CALUNGY1 (2021 file)  CALUNG=CALUNGY2 (2022 file)
Age of Diagnosis	No suffixes <sup>5</sup>	Y1 for 2021  Y2 for 2022	Example: CHDAGED=CHDAGY1 (2021 file) CHOLAGED=CHOLAGY1 (2021 file)  CHDAGED=CHDAGY2 (2022 file) CHOLAGED=CHOLAGY2 (2022 file)
SDOH <sup>6</sup>	No suffixes	1	Example: SDOHELIG=SDOHELIG1 SDAFRDHOME=SDAFRDHOME1

<sup>4</sup> Starting in 2010, variables were added to indicate whether each reported cancer was in remission.

<sup>5</sup> To maintain a previously implemented 8-character naming convention, some variable names had the last character or two dropped in the renaming process.

<sup>6</sup> The SDOH survey was fielded during Panel 26 Round 1 of the MEPS data collection.

## 2.1.2 Constructed Variables for Selection of Group

The following eight variables were constructed and included on the file to facilitate the selection of appropriate cases for various analyses. Table 2 below contains descriptive statistics for these variables.

YEARIND	1=both years, 2=in 2021 only, and 3=in 2022 only
ALL5RDS	In scope and data collected in all 5 rounds (0=no, 1=yes)
DIED	Died during the two-year survey period (0=no, 1=yes)
INST	Institutionalized for some time during the two-year survey period (0=no, 1=yes)
MILITARY	Active duty military for some time during the two-year survey period (0=no, 1=yes)
ENTRSRVY	Entered survey after beginning of panel (mainly births; also includes persons who had no initial chance of selection who moved into a MEPS sample household) (0=no, 1=yes)
LEFTUS	Moved out of the country after beginning of panel (0=no, 1=yes)
OTHER	Not identified in any of the above analytic groups (0=no, 1=yes)

Table 2. Frequencies and Percentage for Constructed Variables

Variable	Number of Records	Percentage of Records (N=6,078)
YEARIND=1 (i.e., person in both years)	6,579	97.60
ALL5RDS=1 (yes)	6,295	93.38
DIED=1 (yes)	155	2.30
INST=1 (yes)	21	0.31
MILITARY=1 (yes)	11	0.16
ENTRSRVY=1 (yes)	236	3.50
LEFTUS=1 (yes)	13	0.19
OTHER=1 (yes)	20	0.30

Following are examples of situations where these variables would be useful in selecting records for analysis:

- Analysts interested in working only with persons who were in-scope and had data for all five rounds of the panel should subset to cases where ALL5RDS=1.
- If a researcher wanted to include persons who were in-scope and had data for all five rounds of the panel as well as those in the survey at the beginning of the panel who subsequently died, then they would include cases where ALL5RDS=1 or (ENTRSRVY=0 and DIED=1).
- If a researcher wanted to include persons who were in-scope and had data for all five rounds of the panel as well as those who died in the second year of the panel, then they would include cases where ALL5RDS=1 or (DIED=1 and YEARIND=1).

### 2.1.3 Estimation Variables

#### ***Longitudinal Estimations for Panel 26***

The Panel 26 Longitudinal PUF contains a weight variable (LONGWT) and variance estimation variables (VARSTR, VARPSU) that should be applied when producing national estimates for longitudinal analyses. For example, LONGWT applied to the 6,295 cases where ALL5RDS=1 produces a weighted population estimate of 311.5 million. This represents an estimate of the number of persons in the civilian noninstitutionalized population for the entire two-year period from 2021-2022. To obtain estimates of variability (such as the standard error of sample estimates or corresponding confidence intervals) for estimates based on MEPS survey data, one needs to take into account the complex sample design of MEPS by specifying the estimation variables including stratum of sample selection (VARSTR), primary sampling unit (VARPSU) and longitudinal weight (LONGWT).

The Panel 26 Longitudinal PUF also contains a longitudinal SAQ weight variable (LSAQWT). This weight variable should be used to perform longitudinal analyses involving any variables from the self-administered questionnaire (SAQ) which was administered to persons age 18 and older in both rounds 2 and 4 of the survey. The variable SAQRDS24 can be used to identify which persons have SAQ data for both versus only one of the two rounds. Table 3 below provides the estimated population size (i.e., the sum of LSAQWT values) for cases with only one round of SAQ data (i.e., SAQRDS24=0) and for cases with both rounds of SAQ data (i.e., SAQRDS24=1). The estimated population size for analyses based on the 3,206 cases with SAQ data for both rounds (i.e., SAQRDS24=1) is 215.8 million.

Table 3. Number of Respondents and Estimated Population Size for SAQ Analyses

Value of SAQRDS24	Description	Number of Respondents (Unweighted)	Estimated Population Size (Weighted by LSAQWT)
0	Persons with one round of SAQ data	3,535	41,954,137
1	Persons with both rounds of SAQ data	3,206	215,829,157
Total	All SAQ respondents	6,741	257,783,294

#### ***Pooled Estimations***

When analyzing subpopulations and/or low-prevalence events, it may be necessary to pool together data from multiple MEPS-HC panels to accumulate a large enough sample size for producing reliable estimates. To ensure accurate variance estimation in such pooled analyses, a consistent and appropriate variance structure must be applied.

MEPS longitudinal weight files for Panels 1–6 were released using panel-specific variance structures. Beginning with Panel 7, however, longitudinal files adopted a common variance structure. This common structure was subsequently revised starting with Panel 24.

To ensure correct variance estimation when pooling longitudinal files, the guidance below should be followed:

1. **Pooling within Panels 7–23 or within Panels 24 and beyond:**  
Simply use the variance strata and PSU variables (**VARSTR**, **VARPSU**)<sup>7</sup> provided on the longitudinal files.
2. **Pooling that involves either:**
  - a) Any panel from Panels 1–6, or
  - b) Any earlier panel in combination with Panels 24 and beyond:  
Use the variance structure from the pooled linkage public use file [HC-036](#), which contains the appropriate consistent variance structure for such combinations.

The [HC-036](#) file is updated annually to include the correct variance structures through the most recent year. Additional information, including a summary chart outlining the appropriate variance structures for various pooling scenarios, can be found in the public use documentation for [HC-036](#) (see Page C-1 for the chart).

---

<sup>7</sup> Note that variable names for strata and PSU are VARSTR and VARPSU, respectively, in longitudinal files for Panel 9 and beyond. These variables were named differently in the longitudinal files for Panel 7 (VARSTRP7, VARPSUP7) and Panel 8 (VARSTRP8, VARPSUP8) and need to be standardized when pooling with subsequent panels.