

MEPS HC-236

Panel 23, 4-Year Longitudinal Data File

December 2023

Due to the COVID-19 pandemic, 2020 data collection moved primarily to phone rather than in-person. This posed a challenge in Panel 25 Round 1, which was difficult to start via phone, resulting in a low response rate. To balance this and increase the number of completes to be comparable to previous years, Panels 23 and 24 were extended to nine rounds of data collection. Phone data collection and the challenges of the pandemic present concerns about data quality. Please take this into consideration when comparing to or pooling with previous years.

**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

<u>Section</u>		<u>Page</u>
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-10
	2.1.3 Estimation Variables	C-11

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. 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 five rounds of interviews covering two full calendar years. Additional rounds were added 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. 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 noninstitutionalized population. In 2006, the NHIS implemented a new sample design, which included Asian persons in addition to households with Black and Hispanic persons in the oversampling of minority populations. NHIS introduced a new sample design in 2016 that discontinued oversampling of these minority groups.

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 cannot 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 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. Data are collected under contract with Westat, Inc. (MEPS HC) and Research Triangle Institute (MEPS MPC). 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 website](#).

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 23, 4-year longitudinal data file from the Medical Expenditure Panel Survey Household Component (MEPS-HC). Released as an ASCII file (with related SAS, STATA, SPSS, and R programming statements and data user information), a SAS data set, a SAS transport dataset, a STATA dataset, and an Excel file, this public use file provides information collected on a nationally representative sample of the civilian noninstitutionalized population of the United States for the four-year period 2018-2021. The file contains 5,334 variables and has a logical record length of 15,157 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 1-9 of MEPS Panel 23 and can be used to analyze changes over a four-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 2018, 2019, 2020, and 2021 Full-Year Consolidated Files (HC-209, HC-216, HC-224, and HC-233, 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 23, 4-year data file is provided in a separate file (H236CB.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

This public use file contains records for 7,080 persons in Panel 23 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 four-year period. Only persons with positive person-level weights (PERWT18F, PERWT19F, PERWT20F, or PERWT21F) are included in the longitudinal PUF data. Data are available for all nine rounds for 87.46% of the cases (6,192 persons). The remaining 12.54% (888 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 four-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 four-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 2018, 2019, 2020, and 2021 Full-Year Consolidated Files (HC-209, HC-216, HC-224, and HC-233, respectively). A variable named YEARIND was added to identify the years of data in which each sample person is included. YEARIND may take the following values for Panel 23:

1 = All four years	9 = 2019 and 2020 only
2 = 2018 only	10 = 2019 and 2021 only
3 = 2019 only	11 = 2020 and 2021 only
4 = 2020 only	12 = 2018, 2019, and 2020 only
5 = 2021 only	13 = 2018, 2019, and 2021 only
6 = 2018 and 2019 only	14 = 2018, 2020, and 2021 only
7 = 2018 and 2020 only	15 = 2019, 2020, and 2021 only
8 = 2018 and 2021 only	

Names for time dependent variables from the full year files were 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.¹ Generally, annual variables with a suffix of “18”, “19”, “20”, and “21” are renamed with a suffix of “Y1”, “Y2”, “Y3”, and “Y4”, 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 2018 full-year file, “3”, “4”, or “5” for variables originating from Rounds 3-5 on the 2019 full-year file, “5”, “6”, or “7” for variables originating from Rounds 5-7 on the 2020 full-year file, and “7”, “8”, or “9” for variables originating from Rounds 7-9 on the 2021 full-year file).² It is necessary to use this crosswalk in conjunction with documentation for the 2018, 2019, 2020, and 2021 full-year consolidated files 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 Files and the Panel 23, 4 Year Longitudinal File

¹ 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).

² While Round 3 values were obtained for most observations from the 2019 Full Year Consolidated File, they were obtained from the 2018 Full Year Consolidated File for sample persons not in the 2019 data (YEARIND=2, 7, 8, or 14). Similarly, values for health insurance and employment-related Round 5 variables were obtained for most observations from the 2020 Full Year Consolidated File, but were obtained from the 2019 Full Year Consolidated File for sample persons not in the 2020 data (YEARIND=3, 6, 10, or 13). Values for all other Round 5 variables were obtained from the 2019 Full Year Consolidated File. Round 7 values were obtained for most observations from the 2021 Full Year Consolidated file but were obtained from the 2020 Full Year Consolidated file for sample persons not in the 2021 data (YEARIND=4, 7, 9, or 12).

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File 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 DOBYY=DOBYY DUID=DUID PID=PID DUPERSID=DUPERSID EDUCYR=EDUCYR HIDEG=HIDEG HISPANX=HISPANX HISPNCAT=HISPNCAT HWELLSPK=HWELLSPK INTVLANG=INTVLANG OTHLGSPK=OTHLGSPK PANEL=PANEL PID=PID RACEAX=RACEAX RACEBX=RACEBX 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 Y3 or YR3 Y4 or YR4	All variables: FAMIDYR=FAMIDYR1 (2018 file) FAMRFPYR=FAMRFPY1 (2018 file) FAMSZEYR=FAMSZYR1 (2018 file) FAMIDYR=FAMIDYR2 (2019 file) FAMRFPYR=FAMRFPY2 (2019 file) FAMSZEYR=FAMSZYR2 (2019 file) FAMIDYR=FAMIDYR3 (2020 file) FAMRFPYR=FAMRFPY3 (2020 file) FAMSZEYR=FAMSZYR3 (2020 file) FAMIDYR=FAMIDYR4 (2021 file) FAMRFPYR=FAMRFPY4 (2021 file) FAMSZEYR=FAMSZYR4 (2021 file)

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Annual, CPS family identifiers	No suffix	Y1 Y2 Y3 Y4	All variables: CPSFAMID= CPSFAMY1 (2018 file) CPSFAMID= CPSFAMY2 (2019 file) CPSFAMID= CPSFAMY3 (2020 file) CPSFAMID= CPSFAMY4 (2021 file)
Annual, health insurance eligibility units	No suffix	Y1 Y2 Y3 Y4	All variables: HIEUIDX=HIEUIDY1 (2018 file) HIEUIDX=HIEUIDY2 (2019 file) HIEUIDX=HIEUIDY3 (2020 file) HIEUIDX=HIEUIDY4 (2021 file)
Annual, in-scope variables	No suffixes	YR1 YR2 YR3 YR4	All variables: INSCOPE=INSCPYP1 (2018 file) INSCOPE=INSCPYP2 (2019 file) INSCOPE=INSCPYP3 (2020 file) INSCOPE=INSCPYP4 (2021 file)
12/31 status variables	1231 in 2018 file 1231 in 2019 file 1231 in 2020 file 1231 in 2021 file	Y1 Y2 Y3 Y4	All variables: FAMS1231=FAMSY1 (2018 file) FCRP1231=FCRPY1 (2018 file) FCSZ1231=FCSZY1 (2018 file) FMRS1231=FMRSY1 (2018 file) INSC1231=INSCY1 (2018 file) FAMS1231=FAMSY2 (2019 file) FCRP1231=FCRPY2 (2019 file) FCSZ1231=FCSZY2 (2019 file) FMRS1231=FMRSY2 (2019 file) INSC1231=INSCY2 (2019 file) FAMS1231=FAMSY3 (2020 file) FCRP1231=FCRPY3 (2020 file) FCSZ1231=FCSZY3 (2020 file) FMRS1231=FMRSY3 (2020 file) INSC1231=INSCY3 (2020 file) FAMS1231=FAMSY4 (2021 file) FCRP1231=FCRPY4 (2021 file) FCSZ1231=FCSZY4 (2021 file) FMRS1231=FMRSY4 (2021 file) INSC1231=INSCY4 (2021 file)

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Annual	18, 18X, 18F, or 18C 19, 19X, 19F, or 19C 20, 20X, 20F, or 20C 21, 21X, 21F, or 21C	Y1, Y1X, Y1F, or Y1C Y2, Y2X, Y2F, or Y2C Y3, Y3X, Y3F, or Y3C Y4, Y4X, Y4F, or Y4C	Examples: TOTEXP18=TOTEXPY1 AGE18X=AGEY1X TOTEXP19=TOTEXPY2 AGE19X=AGEY2X TOTEXP20=TOTEXPY3 AGE20X=AGEY3X TOTEXP21=TOTEXPY4 AGE21X=AGEY4X
Variables for health insurance prior to January 1, 2018 (data collected in Round 1 only)	No suffixes	No suffixes	All variables: PREVCOVR=PREVCOVR MORECOVR=MORECOVR

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Annual	No suffixes ³	Y1 Y2 Y3 Y4	Examples: KEYNESS=KEYNESY1 (2018 file) SAQELIG=SAQELIY1 (2018 file) EVRWRK=EVRWRKY1 (2018 file) EVRETIRE=EVRETIY1 (2018 file) AGELAST=AGELSTY1 (2018 file) DIABDX_M18=DIABDXY1_M18 (2018 file) KEYNESS=KEYNESY2 (2019 file) SAQELIG=SAQELIY2 (2019 file) EVRWRK=EVRWRKY2 (2019 file) EVRETIRE=EVRETIY2 (2019 file) AGELAST=AGELSTY2 (2019 file) DIABDX_M18=DIABDXY2_M18 (2019 file) KEYNESS=KEYNESY3 (2020 file) SAQELIG=SAQELIY3 (2020 file) EVRWRK=EVRWRKY3 (2020 file) EVRETIRE=EVRETIY3 (2020 file) AGELAST=AGELSTY3 (2020 file) DIABDX_M18=DIABDXY3_M18 (2020 file) KEYNESS=KEYNESY4 (2021 file) SAQELIG=SAQELIY4 (2021 file) EVRWRK=EVRWRKY4 (2021 file) EVRETIRE=EVRETIY4 (2021 file) AGELAST=AGELSTY4 (2021 file) DIABDX_M18=DIABDXY4_M18 (2021 file)
Monthly	2-character month + 18 2-character month + 19 2-character month + 20 2-character month + 21	2-character month + Y1 2-character month + Y2 2-character month + Y3 2-character month + Y4	Example: PRIJA18=PRIJAY1 (2018 file) PRIJA19=PRIJAY2 (2019 file) PRIJA20=PRIJAY3 (2020 file) PRIJA21=PRIJAY4 (2021 file)

³ To maintain a previously-implemented 8-character naming convention, some variable names had the last character or two dropped in the renaming process. A few variables have names longer than 8 characters because they were modified in 2018 and tagged with an ‘_M18’ suffix. These variables were altered in the same fashion they would have been without the _M18 suffix, and the _M18 suffix was retained.

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Round specific variables for health insurance and employment	31, 31X, or 31H in 2018 file	1, 1X, or 1H for 2018	Examples: EMPST31=EMPST1 (2018 file)
	42, 42X, or 42H in 2018 file	2, 2X, or 2H for 2018	EMPST42=EMPST2 (2018 file)
	53, 53X, or 53H in 2018 file	3, 3X, or 3H for 2018	EMPST53=EMPST3 (2018 file if sample person is not in 2019 data)
	31, 31X, or 31H in 2019 file	3, 3X, 3H for 2019	EMPST31= EMPST3 (2019 file if sample person is in 2019 data)
	42, 42X, or 42H in 2019 file	4, 4X, 4H for 2019	EMPST42=EMPST4 (2019 file)
	53, 53X, or 53H in 2019 file	5, 5X, 5H for 2019	EMPST53=EMPST5 (2019 file if sample person is not in 2020 data)
	31, 31X, or 31H in 2020 file	5, 5X, 5H for 2020	EMPST31= EMPST5 (2020 file if sample person is in 2020 data) ⁴
	42, 42X, or 42H in 2020 file	6, 6X, 6H for 2020	EMPST42=EMPST6 (2020 file)
	53, 53X, or 53H in 2020 file	7, 7X, 7H for 2020	EMPST53=EMPST7 (2020 file if sample person is not in 2021 data)
	31, 31X, or 31H in 2021 file	7, 7X, 7H for 2021	EMPST31= EMPST7 (2021 file if sample person is in 2021 data)
	42, 42X, or 42H in 2021 file	8, 8X, 8H for 2021	EMPST42=EMPST8 (2021 file)
	53, 53X, or 53H in 2021 file	9, 9X, 9H for 2021	EMPST53=EMPST9 (2021 file)

⁴ Using responses in the Round 6 interview, these variables were constructed from status on the date of the Round 5 interview or between the Round 4 and Round 5 interview dates, not the end of the Round 5 reference period, which was typically 12/31/2019.

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
All other round specific variables	31, 31X, or 31H in 2018 file	1, 1X, or 1H for 2018	Examples: RTHLTH31=RTHLTH1 (2018 file)
	42, 42X, or 42H in 2018 file	2, 2X, or 2H for 2018	RTHLTH42=RTHLTH2 (2018 file)
	53, 53X, or 53H in 2018 file	3, 3X, or 3H for 2018	RTHLTH53=RTHLTH3 (2018 file sample person is not in 2019 data)
	31_M18 in 2018 file	1_M18 for 2018	JTPAIN31_M18=JTPAIN1_M18
	42_M18 in 2018 file	2_M18 for 2018	PROVTY42_M18=PROVTY2_M18
	31, 31X, or 31H in 2019 file	3, 3X, 3H for 2019	RTHLTH31= RTHLTH3 (2019 file if sample person is in 2019 data)
	42, 42X, or 42H in 2019 file	4, 4X, 4H for 2019	RTHLTH42=RTHLTH4 (2019 file)
	53, 53X, or 53H in 2019 file	5, 5X, 5H for 2019	RTHLTH53=RTHLTH5 (2019 file if sample person is in 2019 data; otherwise -1) ⁵
	31_M18 in 2019 file	3_M18 for 2019	JTPAIN31_M18=JTPAIN3_M18
	42_M18 in 2019 file	4_M18 for 2019	PROVTY42_M18=PROVTY4_M18
	42, 42X, or 42H in 2020 file	6, 6X, 6H for 2020	RTHLTH42=RTHLTH6 (2020 file)
	53, 53X, or 53H in 2020 file	7, 7X, 7H for 2020	RTHLTH53=RTHLTH7 (2020 file if sample person is not in 2021 data)
	42_M18 in 2020 file	6_M18 for 2020	PROVTY42_M18=PROVTY6_M18
	31_M18 in 2020 file	5_M18 for 2020	JTPAIN31_M18=JTPAIN5_M18
53_M18 in 2020 file	7_M18 for 2020	JTPAIN53_M18=JTPAIN7_M18	
31, 31X, or 31H in 2021 file	7, 7X, 7H for 2021	RTHLTH31= RTHLTH7 (2021 file if sample person is in 2021 data)	
42, 42X, or 42H in 2021 file	8, 8X, 8H for 2021	RTHLTH42=RTHLTH8 (2021 file)	
53, 53X, or 53H in 2021 file	9, 9X, 9H for 2021	RTHLTH53=RTHLTH9	
42_M18 in 2021 file	8_M18 for 2021	PROVTY42_M18=PROVTY8_M18	

⁵ As with Panels 1 through 22, Round 5 of Panel 23 collected information up to the end of the calendar year, which was 2019 for Panel 23. Therefore, any sample members with 2020 data but not 2019 data (YEARIND=4, 7, 11, or 14) have Round 5 variables set to -1.

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Diabetes preventive care	1753, 1853, and 1953 in 2018 file 1853, 1953, and 2053 in 2019 file 1953, 2053, and 2153 in 2020 file 2053, 2153, and 2253 in 2021 file	Y0R3 for 2017 Y1R3 for 2018 Y2R3 for 2019 Y1R5 for 2018 Y2R5 for 2019 Y3R5 for 2020 Y2R7 for 2019 Y3R7 for 2020 Y4R7 for 2021 Y3R9 for 2020 Y4R9 for 2021 Y5R9 for 2022	Example: DSEY1753=DSEYY0R3 (2018 file) DSEY1853=DSEYY1R3 (2018 file) DSEY1953=DSEYY2R3 (2018 file) DSEY1853=DSEYY1R5 (2019 file) DSEY1953=DSEYY2R5 (2019 file) DSEY2053=DSEYY3R5 (2019 file) DSEY1953=DSEYY2R7 (2020 file) DSEY2053=DSEYY3R7 (2020 file) DSEY2153=DSEYY4R7 (2020 file) DSEY2053=DSEYY3R9 (2021 file) DSEY2153=DSEYY4R9 (2021 file) DSEY2253=DSEYY5R9 (2021 file)
Job Change	3142 or 4253	12 for 2018 23 for 2018 34 for 2019 45 for 2019 56 for 2020 67 for 2020 78 for 2021 89 for 2021	All variables: CHGJ3142=CHGJ12 (2018 file) CHGJ4253=CHGJ23 (2018 file) YCHJ3142=YCHJ12 (2018 file) YCHJ4253=YCHJ23 (2018 file) CHGJ3142=CHGJ34 (2019 file) CHGJ4253=CHGJ45 (2019 file) YCHJ3142=YCHJ34 (2019 file) YCHJ4253=YCHJ45 (2019 file) CHGJ3142=CHGJ56 (2020 file) CHGJ4253=CHGJ56 (2020 file) YCHJ3142=YCHJ67 (2020 file) YCHJ4253=YCHJ67 (2020 file) CHGJ3142=CHGJ78 (2021 file) CHGJ4253=CHGJ78 (2021 file) YCHJ3142=YCHJ89 (2021 file) YCHJ4253=YCHJ89 (2021 file)
Cancer	No suffixes ⁶	Y1 for 2018 Y2 for 2019 Y3 for 2020 Y4 for 2021	Example: CALUNG=CALUNGY1 (2018 file) CALUNG=CALUNGY2 (2019 file) CALUNG=CALUNGY3 (2020 file) CALUNG=CALUNGY4 (2021 file)

⁶ To maintain a previously implemented 8-character naming convention, some variable names had the last character or two dropped in the renaming process.

Type of Variable	Full-Year Consolidated File Variable Name Suffix	Longitudinal File Variable Name Suffix	Specific cases or examples
Age of Diagnosis	No suffixes ⁶	Y1 for 2018 Y2 for 2019 Y3 for 2020 Y4 for 2021	Example: CHDAGED=CHDAGY1 (2018 file) CHOLAGED=CHOLAGY1(2018 file) CHDAGED=CHDAGY2 (2019 file) CHOLAGED=CHOLAGY2(2019 file) CHDAGED=CHDAGY3 (2020 file) CHOLAGED=CHOLAGY3(2020 file) CHDAGED=CHDAGY4 (2021 file) CHOLAGED=CHOLAGY4(2021 file)

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=All four years, 2=2018 only, 3=2019 only, 4=2020 only, 5=2021 only, 6=2018 and 2019 only, 7=2018 and 2020 only, 8=2018 and 2021 only, 9=2019 and 2020 only, 10=2019 and 2021 only, 11= 2020 and 2021 only, 12=2018, 2019, and 2020 only, 13=2018, 2019, and 2021 only, 14=2018, 2020, and 2021 only, 15=2019, 2020, and 2021 only
- ALL9RDS In scope and data collected in all nine rounds (0=no, 1=yes)
- DIED Died during the four-year survey period (0=no, 1=yes)
- INST Institutionalized for some time during the four-year survey period (0=no, 1=yes)
- MILITARY Active duty military for some time during the four-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=7,080)
YEARIND=1 (i.e., person in all four years)	6,518	92.06
ALL9RDS=1 (yes)	6,192	87.46
DIED=1 (yes)	368	5.20
INST=1 (yes)	64	0.90
MILITARY=1 (yes)	32	0.45
ENTRSRVY=1 (yes)	344	4.86
LEFTUS=1 (yes)	35	0.49
OTHER=1 (yes)	73	1.03

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 nine rounds of the panel should subset to cases where ALL9RDS=1.
- If a researcher wanted to include persons who were in-scope and had data for all nine 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 ALL9RDS=1 or (ENTRSRVY=0 and DIED=1).
- If a researcher wanted to include persons who were in-scope and had data for all nine rounds of the panel, as well as those who were in-scope across all four years but died in the fourth year of the panel, then they would include cases where ALL9RDS=1 or (DIED=1 and YEARIND=1).

2.1.3 Estimation Variables

Longitudinal Estimations for Panel 23

The file 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 6,192 cases where ALL9RDS=1 produces a weighted population estimate of 301.9 million. This represents an estimate of the number of persons in the civilian noninstitutionalized population for the entire four-year period from 2018-2021. 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).

This longitudinal file 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 Rounds 2, 4, 6, and 8 of the survey. The variable SAQRDS2468 can be used to identify which

persons have SAQ data for all four rounds. Similarly, the variable SAQRDS246 can be used to identify which persons have SAQ data for Rounds 2, 4, and 6; the variable SAQRDS468 can be used to identify which persons have SAQ data for Rounds 4, 6, and 8; the variable SAQRDS24 can be used to identify which persons have SAQ data for Rounds 2 and 4; the variable SAQRDS46 can be used to identify which persons have SAQ data for Rounds 4 and 6; and the variable SAQRDS68 can be used to identify which persons have SAQ data for Rounds 6 and 8. Table 3 below provides the estimated population size (i.e., the sum of LSAQWT values) for cases with all four rounds of SAQ data (i.e., SAQRDS2468=1) and for cases with two or three rounds of SAQ data. The estimated population size for analyses based on the 3,069 cases with SAQ data for all four rounds (i.e., SAQRDS2468=1) is 203.5 million.

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

SAQ Variable	Value	Description	Number of Respondents (Unweighted)	Estimated Population Size (Weighted by LSAQWT)
Total	Total	All SAQ respondents	7,080	256,482,907
SAQRDS2468	0	Persons with less than four rounds of SAQ data	4,011	52,985,430
SAQRDS2468	1	Persons with all four rounds of SAQ data	3,069	203,497,477
SAQRDS246	0	Persons without Rounds 2, 4, and 6 of SAQ data	3,363	51,976,881
SAQRDS246	1	Persons with Rounds 2, 4, and 6 of SAQ data	3,717	204,506,026
SAQRDS468	0	Persons without Rounds 4, 6, and 8 of SAQ data	3,893	42,964,648
SAQRDS468	1	Persons with Rounds 4, 6, and 8 of SAQ data	3,187	213,518,259
SAQRDS24	0	Persons without Rounds 2 and 4 of SAQ data	2,550	32,106,987
SAQRDS24	1	Persons with Rounds 2 and 4 of SAQ data	4,530	224,375,920
SAQRDS46	0	Persons without Rounds 4 and 6 of SAQ data	3,175	41,861,176
SAQRDS46	1	Persons with Rounds 4 and 6 of SAQ data	3,905	214,621,731
SAQRDS68	0	Persons without Rounds 6 and 8 of SAQ data	3,745	32,410,162
SAQRDS68	1	Persons with Rounds 6 and 8 of SAQ data	3,335	224,072,744

Pooled Estimations

When analyzing subpopulations and/or low prevalence events, it may be desirable to pool together more than one panel of MEPS-HC data to yield sample sizes large enough to generate reliable estimates. Panel 23 is the first panel to include four years of data, so this four-year file should not be combined with other panels until the four-year Panel 24 longitudinal file is available in late 2024. However, the two-year Panel 23 longitudinal file (HC-217) may be pooled with other two-year longitudinal data files, and the three-year Panel 23 longitudinal file (HC-226) may be pooled with the three-year Panel 24 longitudinal file (HC-235). Please refer to [HC-217 on the MEPS website](#) for information about the two-year Panel 23 longitudinal data file and pooling it with other years, and refer to [HC-226 on the MEPS website](#) for information about the three-year Panel 23 longitudinal file.