

MEPS HC-220C: 2020 Other Medical Expenses

August 2022

Due to the COVID-19 pandemic, changes were made to the 2020 MEPS data collection that analysts should keep in mind when doing trend analysis and pooling years of data. 1) The MEPS moved primarily to a phone rather than in-person survey. 2) Panels 23 and 24 were extended to nine rounds (four years) of data collection as opposed to the historical five rounds (two years). Because of the unforeseeable nature of the pandemic, data collection for 2020 included Round 5 interviews for Panel 23 that were fielded under the assumption that that interview would be the panel's last interview. Researchers using variables related to the first interview of the calendar year should read the documentation for their specific variables to understand the sources of the values for Panel 23.

**Agency for Healthcare Research and Quality
Center for Financing, Access, and Cost Trends
5600 Fishers Lane
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-3
2.1 Codebook Structure	C-3
2.2 Reserved Codes	C-4
2.3 Codebook Format	C-5
2.4 Variable Source and Naming Conventions	C-5
2.4.1 Variable - Source Crosswalk.....	C-5
2.4.2 Expenditure and Source of Payment Variables	C-6
2.5 File Contents.....	C-7
2.5.1 Survey Administration Variables.....	C-7
2.5.2 Other Medical Type Variable (OMTYPE_M18).....	C-8
2.5.3 Condition Codes.....	C-8
2.5.4 Expenditure Data.....	C-8
3.0 Survey Sample Information	C-13
3.1 Discussion of Pandemic Effects on Quality of 2020 MEPS Data	C-13
3.1.1 Summary	C-13
3.1.2 Overview	C-13
3.1.3 The Impact of the Pandemic on some Major Federal Surveys	C-13
3.1.4 Modifications to the MEPS-HC 2020 Sample Design and Implementation Effort in Response to the Pandemic.....	C-14

<u>Section</u>	<u>Page</u>	
3.1.5	Data Quality Issues for MEPS for FY 2020.....	C-14
3.1.6	Discussion and Guidance	C-15
3.2	Sample Weight (PERWT20F).....	C-15
3.3	Details on Person Weight Construction	C-16
3.3.1	MEPS Panel 23 Weight Development Process	C-16
3.3.2	MEPS Panel 24 Weight Development Process	C-18
3.3.3	MEPS Panel 25 Weight Development Process	C-18
3.3.4	The Final Weight for 2020.....	C-19
3.4	Coverage.....	C-19
3.5	Using MEPS Data for Trend Analysis	C-20
4.0	Strategies for Estimation.....	C-21
4.1	Basic Estimates of Utilization, Expenditures, and Sources of Payment	C-21
4.1.1	Type of Records on File (OMTYPE_M18).....	C-21
4.2	Variables with Missing Values.....	C-23
4.3	Variance Estimation (VARPSU, VARSTR).....	C-23
4.3.1	Taylor-series Linearization Method.....	C-23
4.3.2	Balanced Repeated Replication (BRR) Method.....	C-25
5.0	Merging/Linking MEPS Data Files	C-26
5.1	Linking to the Person-Level File.....	C-26
5.2	Linking to the Prescribed Medicines File.....	C-27
5.3	Linking to the Medical Conditions File.....	C-27
	References.....	C-28
D.	Variable-Source Crosswalk	D-1

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 noninstitutionalized 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 (and two additional rounds in 2020 covering a third year to compensate for the smaller number of completed interviews in Panel 25), 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 (NCHS). 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 visits, 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 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).

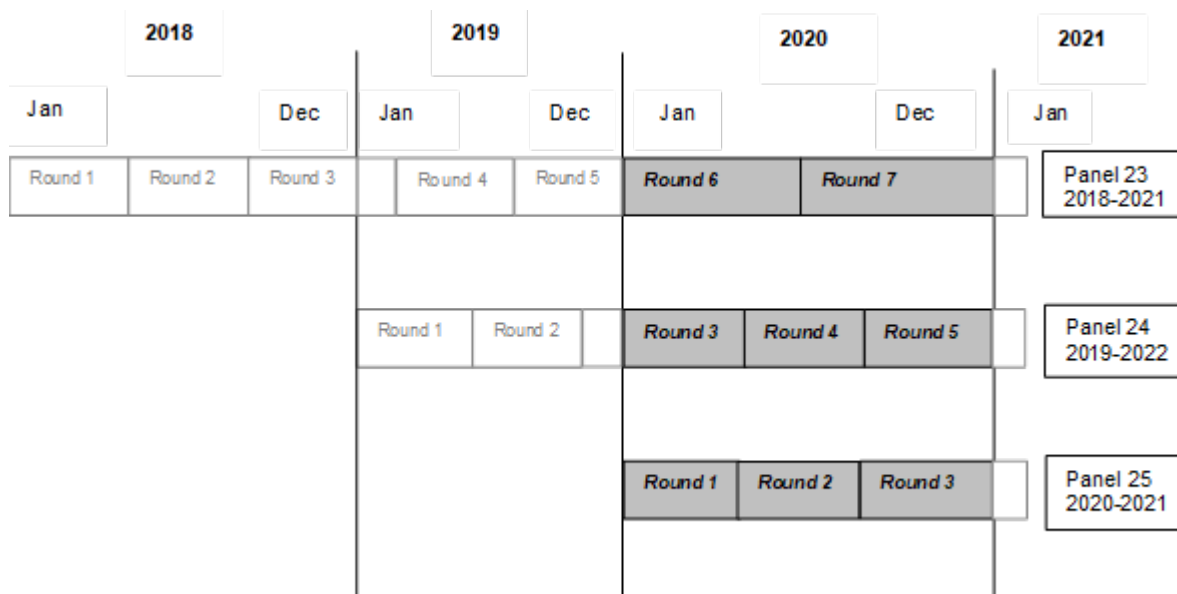
C. Technical and Programming Information

1.0 General Information

This documentation describes one in a series of public use event files from the 2020 Medical Expenditure Panel Survey (MEPS) Household Component (HC). Released as an ASCII data file (with related SAS, SPSS, R, and Stata programming statements and data user information) and a SAS data set, a SAS transport file, a Stata data set, and an Excel file, the 2020 Other Medical Expenses public use event file provides information on the purchases of and expenditures for visual aids, medical equipment, supplies, and other medical items for a nationally representative sample of the civilian noninstitutionalized population of the United States. Data from the Other Medical event file can be used to make estimates of the Other Medical event expenditures associated with medical items for calendar year 2020. The purchase of medical equipment, supplies, and other medical items is based entirely on household reports. They were not included in the Medical Provider Component (MPC); therefore, all expenditure and payment data on the Other Medical event file are reported by the household.

This file contains 23 variables and has a logical record length of 151 with an additional 2-byte carriage return/line feed at the end of each record. As illustrated below, this file consists of MEPS survey data obtained in Round 6 and the 2020 portion of Round 7 for Panel 23; the 2020 portion of Rounds 3 and 5, and all of Round 4 for Panel 24; and Rounds 1, 2, and the 2020 portion of Round 3 of Panel 25 (i.e., the rounds for the MEPS panels covering calendar year 2020).

Full year (FY) 2020 is the first data year to include three panels of data; Panel 23 was extended to include Rounds 6 and 7.



The Other Medical event file contains one record for each type of medical item reported as being purchased or otherwise obtained by a household member during the specified reference period. It should be noted that prior to Panel 21 Round 5 and Panel 22 Round 3, questions regarding glasses/contact lenses were asked in every round, and questions regarding other medical types were asked only in Round 3 and Round 5 for events incurred for the whole year. Starting in Panel 21 Round 5 and Panel 22 Round 3, all OM type questions are asked in every round.

Other medical expense categories included in this file are:

- Glasses or contact lenses
- Ambulance services
- Disposable supplies
- Long-term medical equipment (such as mobility aids, hearing devices, equipment used at home and alterations/modifications, etc.)

Data from this event file can be merged with other 2020 MEPS HC data files for the purpose of appending person-level data, such as demographic characteristics or health insurance coverage, to each Other Medical record.

This file can also be used to construct summary variables of expenditures, source of payment, and related aspects of the purchase of medical items. Aggregate annual person-level information on expenditures for Other Medical equipment is provided on the MEPS 2020 Full Year Consolidated Data File where each record represents a MEPS sampled person.

This document offers a brief overview of the types and levels of data provided, and the content and structure of the file and the codebook. It contains the following sections:

- Data File Information
- Survey Sample Information
- Strategies for Estimation
- Merging/Linking MEPS Data Files
- References
- Variable - Source Crosswalk

For more information on the MEPS HC sample design, see Chowdhury et al (2019). A copy of the MEPS HC survey instrument used to collect the information on the Other Medical expenses file is available on the [MEPS website](#).

2.0 Data File Information

The 2020 Other Medical Expenses public use data set consists of one event-level data file. The file contains characteristics associated with the Other Medical event and imputed expenditure data.

The 2020 Other Medical Expenses public use data set contains 11,970 other medical (OM) expenditure records; of these records, 11,779 are associated with persons having a positive person-level weight (PERWT20F). This file includes records for all household members who resided in eligible responding households and were reported to have purchased or otherwise obtained at least one type of medical item such as medical equipment, glasses, ambulance services, etc. during calendar year 2020. Some persons may have been reported to have obtained more than one type of medical item and, therefore, have several records on this file. On the other hand, persons who were not reported to have obtained a medical item in 2020 have no records on this file. These data were collected during Round 6 and the 2020 portion of Round 7 for Panel 23; the 2020 portion of Rounds 3 and 5, and all of Round 4 for Panel 24; as well as Rounds 1, 2, and the 2020 portion of Round 3 for Panel 25 of the MEPS HC. The persons represented on this file had to meet either (a) or (b) below:

- a) Be classified as a key in-scope person who responded for his or her entire period of 2020 eligibility (i.e., persons with a positive 2020 full-year person-level weight (PERWT20F > 0)), or
- b) Be an eligible member of a family all of whose key in-scope members have a positive person-level weight (PERWT20F > 0). (Such a family consists of all persons with the same value for FAMIDYR.) That is, the person must have a positive full-year family-level weight (FAMWT20F > 0). Note that FAMIDYR and FAMWT20F are variables on the 2020 Full Year Consolidated Data file.

Persons with no Other Medical events for 2020 are not included on this event-level OM file but are represented on the person-level 2020 Full Year Population Characteristics file.

Each record includes the following: type of medical item obtained, imputed sources of payment, total payment and total charge for the medical item, and a full-year person-level weight.

To append person-level information such as demographic or health insurance coverage to each event record, data from this file can be merged with 2020 MEPS HC person-level data (e.g. Full Year Consolidated or Full Year Population Characteristics files) using the person identifier, DUPERSID. Please see Section 5.0 for details on how to merge MEPS data files.

2.1 Codebook Structure

For most variables on the Other Medical Expenses event file, both weighted and unweighted frequencies are provided in the accompanying codebook. The exceptions to this are weight variables and variance estimation variables. Only unweighted frequencies of these variables are

included in the accompanying codebook file. See the Weights Variables list in Section D, Variable-Source Crosswalk.

The codebook and data file sequence list variables in the following order:

- Unique person identifier
- Unique Other Medical expenses identifier
- Type of Other Medical expenses
- Imputed expenditure variables
- Weight and variance estimation variables

Note that the person identifier is unique within this data year.

2.2 Reserved Codes

The following reserved code values are used:

Value	Definition
-1 INAPPLICABLE	Question was not asked due to skip pattern
-7 REFUSED	Question was asked and respondent refused to answer question
-8 DK	Question was asked and respondent did not know answer or the information could not be ascertained
-15 CANNOT BE COMPUTED	Value cannot be derived from data

The value -15 (CANNOT BE COMPUTED) is assigned to MEPS constructed variables in cases where there is not enough information from the MEPS instrument to calculate the constructed variables. “Not enough information” is often the result of skip patterns in the data or from missing information resulting from MEPS responses of -7 (REFUSED) or -8 (DK). Note that reserved code -8 includes cases where the information from the question was “not ascertained” or where the respondent chose “don’t know”.

Generally, values of -1, -7, -8, and -15 for non-expenditure variables have not been edited on this file. The values of -1 and -15 can be edited by the data users/analysts by following the skip patterns in the [HC survey questionnaire](#) located on the MEPS website.

2.3 Codebook Format

The codebook describes an ASCII data set (although the data are also being provided in a SAS data set, a SAS transport file, a Stata data set, and an Excel file). The following codebook items are provided for each variable:

Identifier	Description
Name	Variable name
Description	Variable descriptor
Format	Number of bytes
Type	Type of data: numeric (indicated by NUM) or character (indicated by CHAR)
Start	Beginning column position of variable in record
End	Ending column position of variable in record

2.4 Variable Source and Naming Conventions

In general, variable names reflect the content of the variable. All imputed/edited variables end with an “X”.

As variable collection, universe, or categories are altered, the variable name will be appended with “_Myy” to indicate in which year the alterations took place. Details about these alterations can be found throughout this document.

2.4.1 Variable - Source Crosswalk

Variables were derived from the HC survey questionnaire or from the CAPI. The source of each variable is identified in Section D, Variable-Source Crosswalk in one of four ways:

1. Variables derived from CAPI or assigned in sampling are so indicated as “CAPI derived” or “Assigned in sampling,” respectively;
2. Variables which come from one or more specific questions have those questionnaire sections and question numbers indicated in the “Source” column; questionnaire sections are identified as:
 - EV - Event Roster section
 - CP - Charge Payment section
3. Variables constructed from multiple questions using complex algorithms are labeled “Constructed” in the “Source” column; and
4. Variables that have been edited or imputed are so indicated.

2.4.2 Expenditure and Source of Payment Variables

The names of the expenditure and source of payment variables follow a standard convention, are seven characters in length, and end in an “X” indicating edited/imputed. Please note that imputed means that a series of logical edits, as well as an imputation process to account for missing data, have been performed on the variable.

The total sum of payments and 10 source of payment variables are named in the following way:

The first two characters indicate the type of event:

IP - inpatient stay	OB - office-based visit
ER - emergency room visit	OP - outpatient visit
HH - home health visit	DV - dental visit
OM - other medical equipment	RX - prescribed medicine

In the case of the source of payment variables, the third and fourth characters indicate:

SF - self or family	OF - other federal government
MR - Medicare	SL - state/local government
MD - Medicaid	WC - Workers’ Compensation
PV - private insurance	OT - other insurance
VA - Veterans Administration/CHAMPVA	TR - TRICARE
	XP - sum of payments

In addition, the total charge variable is indicated by TC in the variable name.

The fifth and sixth characters indicate the year (20). The seventh character, “X”, indicates whether the variable is edited/imputed.

For example, OMSF20X is the edited/imputed amount paid by self or family for 2020 other medical equipment and expenditures.

2.5 File Contents

2.5.1 Survey Administration Variables

Person Identifiers (DUID, PID, DUPERSID)

The definitions of Dwelling Units (DUs) in the MEPS Household Survey are generally consistent with the definitions employed for the National Health Interview Survey (NHIS). The dwelling unit ID (DUID) is a seven-digit number consisting of a 2-digit panel number followed by a five-digit random number assigned after the case was sampled for MEPS. A three-digit person number (PID) uniquely identifies each person within the DU. The ten-character variable DUPERSID uniquely identifies each person represented on the file and is the combination of the variables DUID and PID. IDs begin with the 2-digit panel number.

For detailed information on dwelling units and families, please refer to the documentation for the 2020 Full Year Population Characteristics file.

Record Identifiers (EVNTIDX)

EVNTIDX uniquely identifies each Other Medical expense event (i.e., each record on the OME file) and is the variable required to link Other Medical events to data files containing details on prescribed medicines (MEPS 2020 Prescribed Medicines File). EVNTIDX begins with the 2-digit panel number and ends with the 2-digit event type number. For details on linking, see Section 5.0, or the MEPS 2020 Appendix File, HC-220I.

Round Indicator (EVENTRN)

EVENTRN indicates the round in which the Other Medical event was reported. Each record represents a summary of expenditures for items purchased or otherwise obtained for 2020. Starting with Panel 22, the Round 3 questions relating to the number of times glasses or contact lenses were obtained in each year of a panel are eliminated due to design changes. Instead, if a person's reference period crosses between year one and year two of a panel, the question 'whether a particular OM type was purchased/used' for each of the four OM types (glasses/contacts, ambulance services, disposable supplies and long-term medical equipment) is asked separately for each of two years of a panel. Please note: Rounds 6 and 7 (partial) are associated with MEPS survey data collected from Panel 23. Likewise, Rounds 3 (partial), 4, and 5 (partial) are associated with data collected from Panel 24, and Rounds 1, 2, and 3 (partial) are associated with data collected from Panel 25.

Panel Indicator (PANEL)

PANEL is a constructed variable used to specify the panel number for the person. PANEL will indicate either Panel 23, Panel 24, or Panel 25 for each person on the file. Panel 23 is the panel

that started in 2018, Panel 24 is the panel that started in 2019, and Panel 25 is the panel that started in 2020.

2.5.2 Other Medical Type Variable (OMTYPE_M18)

Other medical expenditures (OMTYPE_M18) include glasses or contact lenses, ambulance services, disposable supplies and long-term medical equipment. The OM type questions are asked in every round.

2.5.3 Condition Codes

Conditions data are not collected for Other Medical events; therefore, this file cannot be linked to the Conditions File.

2.5.4 Expenditure Data

Definition of Expenditures

Expenditures on this file refer to what is paid for the medical item. More specifically, expenditures in MEPS are defined as the sum of payments for each medical item that was obtained, including out-of-pocket payments and payments made by private insurance, Medicaid, Medicare and other sources. The definition of expenditures used in MEPS differs slightly from its predecessors, the 1987 NMES and 1977 NMCES surveys, where “charges” rather than sum of payments were used to measure expenditures. This change was adopted because charges became a less appropriate proxy for medical expenditures during the 1990s due to the increasingly common practice of discounting. Although measuring expenditures as the sum of payments incorporates discounts in the MEPS expenditure estimates, these estimates do not incorporate any payment not directly tied to specific medical care events, such as bonuses or retrospective payment adjustments paid by third party payers. The charges associated with uncollected liability, bad debt, and charitable care (unless provided by a public clinic or hospital) are not counted as expenditures because there are no payments associated with those classifications. While charge data are provided on this file, data users/analysts should use caution when working with these data because a charge does not typically represent actual dollars exchanged for services or the resource costs of those services, nor are they directly comparable to the expenditures defined in the 1987 NMES. For details on expenditure definitions, please refer to the following, “Informing American Health Care Policy” (Monheit et al., 1999). AHRQ has developed factors to apply to the 1987 NMES expenditure data to facilitate longitudinal analysis. These factors can be accessed via the CFACT data center. For more information see the [Data Center section of the MEPS website](#). If examining trends in MEPS expenditures, please refer to Section C, sub-Section 3.5 for more information.

Data Editing and Imputation Methodologies of Expenditure Variables

The general methodology used for editing and imputing expenditure data is described below. The MPC did not include either the dental events or other medical expenditures (such as glasses, ambulance, and disposable supplies). Therefore, although the general procedures remain the same for dental and other medical expenditures, editing and imputation methodologies were applied only to household-reported data. Please see below for details on the differences between these editing/imputation methodologies. Separate imputations were performed for simple events, as well.

General Data Editing Methodology

Logical edits were used to resolve internal inconsistencies and other problems in the HC survey-reported data. The edits were designed to preserve partial payment data from households and providers, and to identify actual and potential sources of payment for each household-reported event. In general, these edits accounted for outliers, copayments or charges reported as total payments. In addition, edits were implemented to correct for misclassifications between Medicare and Medicaid, and between Medicare HMOs and private HMOs as payment sources. These edits produced a complete vector of expenditures for some events, and provided the starting point for imputing missing expenditures in the remaining events.

Imputation Methodologies

The predictive mean matching imputation method was used to impute missing expenditures. This procedure uses regression models (based on events with completely reported expenditure data) to predict total expenses for each event. Then, for each event with missing payment information, a donor event with the closest predicted payment with the same pattern of expected payment sources as the event with the missing payment was used to impute the missing payment value.

A weighted sequential hot-deck procedure was used to impute the missing total charges. This procedure uses survey data from respondents to replace missing data, while taking into account the persons' weighted distribution in the imputation process.

Other Medical Expenses Data Editing and Imputation

The CAPI instrument collects the total charge and out-of-pocket expenditures for disposable supplies (OMTYPE_M18=3) in a range format. The ranges were replaced with mean dollar amounts of respective expenditures reported in each range in prior years. The mean dollar replacement amounts are shown in the tables below:

Total Charge Range for OMTYPE_M18=3	Mean Dollar Amounts
\$0	\$0
\$1 - \$10	\$8.10
\$11 - \$30	\$20.50
\$31 - \$100	\$57.80
\$101 or more	\$1,571.70

Out of Pocket Payment Range for OMTYPE_M18=3	Mean Dollar Amount
\$0	\$0
\$1 - \$10	\$6.70
\$11 - \$30	\$20.40
\$31 - \$100	\$56.20
\$101 or more	\$442.60
-7, -8, -15	-8

Expenditures on other medical equipment and services were developed in a sequence of logical edits and imputations. The household edits were used to correct obvious errors in the reporting of expenditures, and to identify actual and potential sources of payments. Some of the edits were global (i.e., applied to all events). Others were hierarchical and mutually exclusive.

Logical edits were used to sort each event into a specific category for the imputations. Events with complete expenditures were flagged as potential donors for the predictive mean imputations, while events with missing expenditure data were assigned to various recipient categories. Each event with missing expenditure data was assigned to a recipient category based on the extent of its missing charge and expenditure data. For example, an event with a known total charge but no expenditure information was assigned to one category, while an event with a known total charge and partial expenditure information was assigned to a different category. Similarly, events without a known total charge and no or partial expenditure information were assigned to various recipient categories.

The logical edits produced nine recipient categories for events with missing data. Eight of the categories were for events with a common pattern of missing data and a primary payer other than Medicaid. Medicaid events were imputed separately because persons on Medicaid rarely know the provider's charge for services or the amount paid by the state Medicaid program. As a result, the total charge for Medicaid-covered services was imputed and discounted to reflect the amount that a state program might pay for the care.

Separate predictive mean imputations were used to impute missing data in each of the eight recipient categories. The donor pool included "free events" because in some instances, providers are not paid for their services. These events represent charity care, bad debt, provider failure to bill, and third party payer restrictions on reimbursement in certain circumstances. If free events were excluded from the donor pool, total expenditures would be over-counted because the distribution of free events among complete events (donors) is not represented among incomplete events (recipients).

Imputation Flag Variable (IMPFLAG)

IMPFLAG is a six-category variable that indicates if the event contains complete Household Component (HC) or Medical Provider Component (MPC) data, was fully or partially imputed, or

was imputed in the capitated imputation process (for OP and MV events only). The following list identifies how the imputation flag is coded; the categories are mutually exclusive.

IMPFLAG = 0 not eligible for imputation (includes zeroed out events)

IMPFLAG = 1 complete HC data

IMPFLAG = 2 complete MPC data (not applicable to OM events)

IMPFLAG = 3 fully imputed

IMPFLAG = 4 partially imputed

IMPFLAG = 5 complete MPC data through capitation imputation (not applicable to OM events)

Flat Fee Expenditures

A flat fee is the fixed dollar amount a person is charged for a package of health care services provided during a defined period of time. Beginning in 2018, the Other Medical service events can no longer be reported as a flat fee group.

Zero Expenditures

Some respondents reported persons obtaining medical items where the payments were zero. This could occur when (1) the item or service was free or (2) bad debt was incurred. If all of the medical events for a person fell into one of these categories, then the total annual expenditures for that person would be zero.

Sources of Payment

In addition to total expenditures, variables are provided which itemize expenditures according to major source of payment categories. These categories are:

1. Out-of-pocket by User (self or family) - includes any deductible, coinsurance, and copayment amounts not covered by other sources, as well as payments for services and providers not covered by the person's insurance or other sources,
2. Medicare,
3. Medicaid,
4. Private Insurance,
5. Veterans Administration/CHAMPVA, excluding TRICARE,

6. TRICARE,
7. Other Federal Sources - includes Indian Health Service, military treatment facilities, and other care by the federal government,
8. Other State and Local Source - includes community and neighborhood clinics, state and local health departments, and state programs other than Medicaid,
9. Workers' Compensation, and
10. Other Unclassified Sources - includes sources such as automobile, homeowner's, and liability insurance, and other miscellaneous or unknown sources.

Other Medical Expenditure Variables (OMSF20X-OMTC20X)

Other medical expenditure data were obtained only through the Household Component Survey. For cases with missing expenditure data, Other Medical expenditures were imputed using the procedures described above.

OMSF20X - OMOT20X are the 10 sources of payment. OMTC20X is the total charge, and OMXP20X is the sum of the 10 sources of payment for the other medical expenditures. The 10 sources of payment are: self/family (OMSF20X), Medicare (OMMR20X), Medicaid (OMMD20X), private insurance (OMPV20X), Veterans Administration/CHAMPVA (OMVA20X), TRICARE (OMTR20X), other federal sources (OMOF20X), state and local (non-federal) government sources (OMSL20X), Workers' Compensation (OMWC20X), and other insurance (OMOT20X).

Rounding

Expenditure variables on the 2020 Other Medical event file have been rounded to the nearest penny. Person-level expenditure information to be released on the MEPS 2020 Full Year Consolidated File will be rounded to the nearest dollar. It should be noted that using the MEPS event files to create person-level totals will yield slightly different totals than those found on the consolidated file. These differences are due to rounding only. Moreover, in some instances, the number of persons having expenditures on the event files for a particular source of payment may differ from the number of persons with expenditures on the person-level expenditure file for that source of payment. This difference is also an artifact of rounding only.

3.0 Survey Sample Information

3.1 Discussion of Pandemic Effects on Quality of 2020 MEPS Data

3.1.1 Summary

Data collection for in-person sample surveys in 2020 presented real challenges after the onset of the COVID-19 pandemic at a national level in mid-March of that year. After major modifications to the standard MEPS study design, it was possible to collect data safely, but there were naturally concerns about the quality of the data after such modifications. Some issues related to data quality were identified and are discussed below. As with most in-person surveys conducted in 2020, researchers are counseled to take care in the interpretation of 2020 estimates including the comparison of such estimates with those of other years.

3.1.2 Overview

The onset of the COVID-19 pandemic in 2020 had a major impact on the MEPS Household Component (MEPS-HC) as it did for most major federal surveys and, of course, American life generally. The following discussion describes 1) the general impact of the pandemic on three major federal surveys (the effects on two of which also affect MEPS); 2) modifications to the MEPS sample design and field operations in 2020 due to the pandemic; and 3) potential data quality issues in the FY 2020 MEPS data related to the COVID-19 pandemic.

3.1.3 The Impact of the Pandemic on some Major Federal Surveys

Many important federal surveys were collecting data when much of the nation shut down in the face of the pandemic in March 2020. Among them were the Current Population Survey (CPS), the American Community Survey (ACS), and the National Health Interview Survey (NHIS). The ACS and the NHIS field new samples each year. The CPS includes rotating panels, meaning some of the sampled households fielded had participated in prior years while others were fresh. Two of these surveys have important roles in MEPS. Estimates of CPS subgroups serve as benchmarks for the MEPS weighting process (referred to below as “raking control totals”) while households fielded for Round 1 of MEPS in each year are selected as a subsample of the NHIS responding households from the prior year.

Because data collection in 2020 occurred under such unusual circumstances, all three of these surveys have reported bias concerns. (In fact, the ACS decided not to release a standard database for 2020 due to the uncertain quality of the data, while the CPS and the NHIS released data but included reports discussing concerns about bias.) All three surveys have reported evidence of nonresponse bias, specifically, that households in higher socio-economic levels were relatively more likely to respond and the sample weighting was unable to fully compensate for this. As a result, analysts have been cautioned about the accuracy of survey estimates and the ability to compare resulting estimates with estimates obtained in the years prior to the pandemic.

The quality of CPS data is of particular importance to Full Year 2020 MEPS PUFs as CPS estimates serve as the control totals for the raking component of the MEPS weighting process. These control totals are based on the following demographic variables: age, sex, race/ethnicity, region, MSA status, educational attainment, and poverty status. The CPS estimates used in the development of the FY 2020 MEPS PUF weights that were based on the variables age, sex, race/ethnicity, region, and MSA status were evaluated by the Census Bureau and determined to be of high quality. However, similar evaluations of the corresponding CPS estimates associated with educational attainment and poverty status found that these estimates suffered from bias.

A set of references discussing the fielding of these three surveys during the pandemic and resulting bias concerns can be found in the References section of this document.

3.1.4 Modifications to the MEPS-HC 2020 Sample Design and Implementation Effort in Response to the Pandemic

For the MEPS-HC, face-to-face interviewing ceased due to the COVID-19 pandemic on March 17, 2020. At that time, there were two MEPS panels in the field for which 2020 data were being collected: Round 1 of Panel 25 and Round 3 of Panel 24. The sampled households for Panel 25 were being contacted and asked to participate in MEPS for the first time while those from Panel 24 had already participated in MEPS for two rounds. A third MEPS panel was also in the field in early 2020, Round 5 of Panel 23, collecting data for the last portion of 2019.

In developing a plan for how best to resume MEPS data collection, the primary issues were how to do so safely for both sampled household members and interviewers and the potential impact on data quality. Telephone data collection, although not the preferred method of data collection in general for MEPS-HC, was the natural option because it did not require in-person contact with respondents and could be implemented relatively quickly. The impact of changing to telephone on both response rates and data quality was expected to be larger for Panel 25 Round 1 (e.g., no experience with reporting health care events in the recent past). At the time in-person interviewing stopped in mid-March 2020 completion rates for Panels 23 and 24 were substantially higher than those for Panel 25.

AHRQ decided to field Panel 23 for at least one more year, asking Panel 23 respondents if they would be open to further participation in MEPS in newly added Rounds 6 and 7. Extending Panel 23 was meant to both offset the decrease in the number of cases in the FY 2020 data related to lower expected sample yields for Panel 25 and to improve data quality by retaining a set of participants who were familiar with MEPS. These decisions required major changes in survey operations, including adding a fall Panel 23 Round 6 interview covering all 2020 events from January 1, 2020 to the date of the interview.

3.1.5 Data Quality Issues for MEPS for FY 2020

Numerous analyses were conducted to examine potential impacts on data quality and to gain a more complete understanding of these issues. Zuvekas and Kashihara (2021) discuss some of these analyses and provide additional background information on how the MEPS study design was modified in 2020 in response to the pandemic. Three sources of potential bias that were

identified are noted here: the long recall period for Round 6 of Panel 23; switching from in-person to telephone interviewing which likely had a larger impact on Panel 25; and the impact of CPS bias on the MEPS weights. Each is considered in turn.

Comparisons of health care utilization data for Panel 24 and Panel 23 indicated that the extended reference period for Panel 23 Round 6 may have resulted in recall issues for respondents. Round 6 was initially fielded in the late summer and early fall of 2020, and because the Round 5 reference period ended on December 31, 2019, the recall period for health care events and related information extended back to January 1, 2020, much longer than for typical MEPS rounds. For Panel 23 Round 6 respondents, events of a less salient nature, such as dental visits and office-based physician visits, occurring in early 2020 were under-reported. Underreporting was confirmed through both an examination of differential utilization across 2020 for Panel 23 respondents as well as statistical comparisons of Panel 23 and Panel 24 event estimates. Adjustments were made to the sample weights for Panel 23 to help address this concern. Details on these adjustments can be found in Section 3.3.1.

Comparisons of Panel 25 with Panel 24 health care utilization data found that the difference in estimates reached statistical significance for several event types with those from Panel 25 generally being the higher. The same comparisons between first and second year panels in MEPS in recent years showed relatively few such differences with no differences at all in 2019.

Finally, AHRQ decided to calibrate, via raking, the FY 2020 Consolidated PUF weights to control totals reflecting CPS 2021 poverty status data. As discussed earlier, bias was identified by the Census Bureau in the 2020 and 2021 CPS income data and correlates. Nevertheless, the Census Bureau decided to use its standard sample weighting approach for both the 2020 and 2021 CPS ASEC data sets while recognizing some deficiencies in the nonresponse adjustment approach for the two years as a result of data collection during the pandemic. Similarly, MEPS has used poverty status based on the CPS estimates for calibration for many years and continued to do so for the 2020 Full Year Consolidated PUF as it was decided that the advantages of doing so outweighed the disadvantages.

3.1.6 Discussion and Guidance

The additional procedures for developing person-level and family-level final weights for the 2020 Consolidated MEPS data were designed to correct for potential biases in the data due to changes in data collection and response bias. However, evaluations of MEPS data quality in 2020 - corroborated in analyses of other Federal surveys fielded in 2020 - suggest that users of the MEPS FY 2020 Consolidated PUF should exercise caution when interpreting estimates and assessing analyses based on these data as well as in comparing 2020 estimates to those of prior years.

3.2 Sample Weight (PERWT20F)

There is a single full-year person-level weight (PERWT20F) assigned to each record for each key, in-scope person who responded to MEPS for the full period of time that he or she was

in-scope during 2020. A key person was either a member of a responding NHIS household at the time of interview or joined a family associated with such a household after being out-of-scope at the time of the NHIS (the latter circumstance includes newborns as well as those returning from military service, an institution, or residence in a foreign country). A person is in-scope whenever he or she is a member of the civilian noninstitutionalized portion of the U.S. population.

3.3 Details on Person Weight Construction

The person-level weight PERWT20F was developed in several stages. Person-level weights for Panel 23, Panel 24, and Panel 25 were created separately. The weighting process for each panel included an adjustment for nonresponse over time and calibration to independent population figures. The calibration was initially accomplished separately for each panel by raking the corresponding sample weights for those in-scope at the end of the calendar year to Current Population Survey (CPS) population estimates based on six variables. The six variables used in the establishment of the initial person-level control figures were: educational attainment of the reference person (no degree, high school/GED no college, some college, bachelor's degree or higher); census region (Northeast, Midwest, South, West); MSA status (MSA, non-MSA); race/ethnicity (Hispanic; Black, non-Hispanic; Asian, non-Hispanic; and other); sex; and age. A 2020 composite weight was then formed by multiplying each weight from Panel 23 by the factor .29, each weight from Panel 24 by the factor .36, and each weight from Panel 25 by the factor .35. The choice of factors reflected the relative sample sizes of the three panels, helping to limit the variance of estimates obtained from pooling the three samples. The composite weight was raked to the same set of CPS-based control totals.

The standard approach for MEPS weighting is as follows. When the poverty status information derived from income variables becomes available, a final raking is undertaken. The full sample weight appearing on the Population Characteristics PUF for a given year is re-raked, establishing control figures reflecting poverty status rather than educational attainment. Thus, control totals are established using poverty status (five categories: below poverty, from 100 to 125 percent of poverty, from 125 to 200 percent of poverty, from 200 to 400 percent of poverty, at least 400 percent of poverty) as well as the other five variables previously used in the weight calibration.

This approach was modified for the full sample weights appearing on the FY 2020 Consolidated PUF. The raking of the Panel 23 weights was re-done as described in Section 3.3.1 below, and then the resulting Panel 23 weights were composited with those previously established for Panels 24 and 25 with the same factors as described previously, producing a new full sample weight. This new weight was then raked to control figures reflecting the standard five variables plus poverty status.

3.3.1 MEPS Panel 23 Weight Development Process

The person-level weight for MEPS Panel 23 was developed using the 2019 full-year weight for an individual as the initially assigned weight for 2019 survey participants present in 2020. For key, in-scope members who joined an RU some time in 2020 after being out-of-scope in 2019, the initially assigned person-level weight was the corresponding 2019 family weight. The

weighting process included an adjustment for person-level nonresponse over Rounds 6 and 7 as well as raking to population control figures for December 2020 for key, responding persons in-scope on December 31, 2020. These control totals were derived by scaling back the population distribution obtained from the March 2021 CPS to reflect the December 31, 2020 estimated population total (estimated based on Census projections for January 1, 2021). Variables used for person-level raking included: education of the reference person (three categories: no degree; high school/GED only or some college; Bachelor’s or higher degree); Census region (Northeast, Midwest, South, West); MSA status (MSA, non-MSA); race/ethnicity (Hispanic; Black, non-Hispanic; Asian, non-Hispanic; and other); sex; and age. (It may be noted that for confidentiality reasons, the MSA status variables are no longer released for public use. This started with the Full-Year 2013 Person-Level Use PUF.) The final weight for key, responding persons who were not in-scope on December 31, 2020 but were in-scope earlier in the year was the nonresponse-adjusted person weight without raking.

In developing the person-level weight for Panel 23, an additional raking dimension was included beyond those based on the usual six variables. This dimension was added to adjust the distribution of event-based (i.e., office-based [MV] and/or outpatient [OP]) estimates to align with corresponding Panel 24 weighted estimates. The table below shows ratios of weighted totals (population estimates) associated with this additional raking dimension, reflecting the extent to which the Panel 23 estimates were modified in order to correspond to Panel 24 estimates. Generally, the weights of the records with any event in Q1 are inflated to account for the under reporting of events in Q1.

Ratio of Adjusted to Unadjusted Weights

# of Events	Ratio
1: No MV/OP Events	0.8375
2: At least 1 event in Q1 and no events in other quarters	2.7509
3: At least 1 event in Q2 and no events in other quarters	0.9456
4: At least 1 event in Q3 and no events in other quarters	0.7811
5: At least 1 event in Q4 and no events in other quarters	0.7149
6: At least 1 event in Q1 and at least 1 event in at least 1 other quarter	1.3188
7: At least 1 event in Q2 and at least 1 event in at least 1 Q3 or Q4	0.7199
8: Other	0.6908

The Panel 23 2019 full-year weight used as the base weight for Panel 23 was derived from the 2018 MEPS Round 1 weight and reflected adjustment for nonresponse over the remaining data collection rounds in 2018 and 2019 as well as raking to the December 2018 and December 2019 population control figures.

For the raking variable “education of the reference person” there were four raking categories in prior years: no degree; high school/GED no college; some college; and Bachelor's or a higher degree. However, as mentioned in the discussion of data quality issues in 2020 in Section 3.1,

there was evidence that the onset of the COVID-19 pandemic in the years of 2020 and 2021 affected estimates associated with income and education (further details can be found in the references associated with the CPS data quality issues in 2020 and 2021 in the References section). For the full-year 2019 weights, March 2019 CPS was utilized instead of March 2020 CPS in the construction of control totals to avoid data quality issues connected to the COVID-19 pandemic. For the full-year 2020 weights, since there are no reliable education estimates from 2020 or 2021 CPS, a regression approach was implemented to derive education control figures. The regression approach involved two steps. The first step fit a linear regression model for each of the four education categories using the 2013-2018 CPS education of reference person distributions as the predictors in order to estimate the distribution for 2020, and the second step derived the education of reference person control figures by applying the estimated 2020 education distribution to the December 31, 2020 population total. The models for “no degree” and “Bachelor's or a higher degree” performed extremely well with R^2 values of 0.97 and 0.98, respectively. The models for “high school/GED no college” and “some college” showed a lower goodness of fit, especially for some college, with a R^2 value of 0.74. A linear regression for the two categories combined improved the R^2 value to 0.89, so the two levels were combined for the 2020 weight development.

3.3.2 MEPS Panel 24 Weight Development Process

The person-level weight for MEPS Panel 24 was developed using the 2019 full-year weight for an individual as a “base” weight for survey participants present in 2019. For key, in-scope members who joined an RU some time in 2020 after being out-of-scope in 2019, the initially assigned person-level weight was the corresponding 2019 family weight. The weighting process included an adjustment for person-level nonresponse over Rounds 4 and 5 as well as raking to population control totals for December 2020 used for the MEPS Panel 23 weights for key, responding persons in-scope on December 31, 2020. The six standard variables employed for Panel 23 raking (education level, census region, MSA status, race/ethnicity, sex, and age) were also used for Panel 24 raking. Similar to Panel 23, the Panel 24 final weight for key, responding persons not in-scope on December 31, 2020 but in-scope earlier in the year was the nonresponse-adjusted person weight without raking.

Note that the 2019 full-year weight that was used as the base weight for Panel 24 was derived as follows; adjustment of the 2019 MEPS Round 1 weight for nonresponse over the remaining data collection rounds in 2019; and raking the resulting nonresponse adjusted weight to December 2019 population control figures.

3.3.3 MEPS Panel 25 Weight Development Process

The person-level weight for MEPS Panel 25 was developed using the 2020 MEPS Round 1 person-level weight as a “base” weight. The MEPS Round 1 weights incorporated the following components: the original household probability of selection for the NHIS, use of a subsample of the NHIS panels and quarters reserved for MEPS, an adjustment for NHIS nonresponse, the probability of selection for MEPS from NHIS responding households, adjustment for nonresponse at the dwelling unit level for Round 1, and poststratification to control figures at the

person level obtained from the March CPS of the corresponding year. For key, in-scope members who joined an RU after Round 1, the Round 1 family weight served as a “base” weight.

The weighting process also included an adjustment for nonresponse over the remaining data collection rounds in 2020 as well as raking to the same population control figures for December 2020 used for the MEPS Panel 23 and Panel 24 weights for key, responding persons in-scope on December 31, 2020. The six standard variables employed for Panel 23 and Panel 24 raking (educational attainment of the reference person, census region, MSA status, race/ethnicity, sex, and age) were also used for Panel 25. The event-based raking dimension used for Panel 23 was not employed for Panel 25. Similar to Panel 23 and Panel 24, the Panel 25 final weight for key, responding persons who were not in-scope on December 31, 2020 but were in-scope earlier in the year was the person weight after the nonresponse adjustment.

3.3.4 The Final Weight for 2020

The final raking of those in-scope at the end of the year has been described above. In addition, the composite weights of three groups of persons who were out-of-scope on December 31, 2020 were adjusted for expected undercoverage. Specifically, the weights of those who were in-scope some time during the year, out-of-scope on December 31, and entered a nursing home during the year and still residing in a nursing home at the end of the year were poststratified to an estimate of the number of persons who were residents of Medicare- and Medicaid-certified nursing homes for part of the year (approximately 3-9 months) during 2014. This estimate was developed from data on the Minimum Data Set (MDS) of the Center for Medicare and Medicaid Services (CMS). The weights of persons who died while in-scope were poststratified to corresponding estimates derived using data obtained from the Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Underlying Cause of Death, 1999-2020 on [CDC WONDER Online Database](#), released in 2022, the latest available data at the time. Separate decedent control totals were developed for the “65 and older” and “under 65” civilian noninstitutionalized populations.

Overall, the weighted population estimate for the civilian noninstitutionalized population for December 31, 2020 is 324,539,180 (PERWT20F >0 and INSC1231=1). The sum of person-level weights across all persons assigned a positive person-level weight is 328,545,297.

3.4 Coverage

The target population for MEPS in this file is the 2020 U.S. civilian noninstitutionalized population. However, the MEPS sampled households are a subsample of the NHIS households interviewed in 2017 (Panel 23), 2018 (Panel 24), and 2019 (Panel 25). New households created after the NHIS interviews for the respective panels and consisting exclusively of persons who entered the target population after 2017 (Panel 23), after 2018 (Panel 24), or after 2019 (Panel 25) are not covered by MEPS. Neither are previously out-of-scope persons who join an existing household but are unrelated to the current household residents. Persons not covered by a given MEPS panel thus include some members of the following groups: immigrants; persons leaving the military; U.S. citizens returning from residence in another country; and persons leaving

institutions. The set of uncovered persons constitutes a relatively small segment of the MEPS target population.

3.5 Using MEPS Data for Trend Analysis

First, of course, we note that there are uncertainties associated with 2020 data quality as discussed in Section 3.1. Evaluations described in that section suggest that care should be taken in the interpretation of estimates based on data collected in 2020 as well as in comparisons over time. Trend analyses are challenging since the advent of the COVID-19 pandemic resulted in uncertain data quality for MEPS as well as standard benchmark sources such as the CPS, ACS, and NHIS while the pandemic also had an impact on the health and access to health care of the U.S. population. For such reasons, the extent to which 2020 health care parameters may differ from those of prior years is difficult to assess.

In terms of other factors to be aware of, MEPS began in 1996, and the utility of the survey for analyzing health care trends expands with each additional year of data; however, it is important to consider a variety of factors when examining trends over time using MEPS. Tests of statistical significance should be conducted to assess the likelihood that observed trends are not attributable to sampling variation. The length of time being analyzed should also be considered. In particular, large shifts in survey estimates over short periods of time (e.g. from one year to the next) that are statistically significant should be interpreted with caution unless they are attributable to known factors such as changes in public policy, economic conditions, or MEPS survey methodology.

With respect to methodological considerations, in 2013 MEPS introduced an effort focused on field procedure changes such as interviewer training to obtain more complete information about health care utilization from MEPS respondents with full implementation in 2014. This effort likely resulted in improved data quality and a reduction in underreporting starting in the second half of 2013 and throughout 2014 full year files and have had some impact on analyses involving trends in utilization across years. The aforementioned changes in the NHIS sample design in 2016 could also potentially affect trend analyses. The new NHIS sample design is based on more up-to-date information related to the distribution of housing units across the U.S. As a result, it can be expected to better cover the full U.S. civilian, noninstitutionalized population, the target population for MEPS as well as many of its subpopulations. Better coverage of the target population helps to reduce the potential for bias in both NHIS and MEPS estimates.

Another change with the potential to affect trend analyses involved modifications to the MEPS instrument design and data collection process, particularly in the events sections of the instrument. These were introduced in the Spring of 2018 and thus affected data beginning with Round 1 of Panel 23, Round 3 of Panel 22, and Round 5 of Panel 21. Since the Full Year 2017 PUFs were established from data collected in Rounds 1-3 of Panel 22 and Rounds 3-5 of Panel 21, they reflected two different instrument designs. In order to mitigate the effect of such differences within the same full year file, the Panel 22 Round 3 data and the Panel 21 Round 5 data were transformed to make them as consistent as possible with data collected under the previous design. The changes in the instrument were designed to make the data collection effort more efficient and easy to administer. In addition, expectations were that data on some items, such as those related to health care events, would be more complete with the potential of

identifying more events. Increases in service use reported since the implementation of these changes are consistent with these expectations. **Data users should be aware of possible impacts on the data and especially trend analyses for these data years due to the design transition.**

Process changes, such as data editing and imputation, may also affect trend analyses. For example, users should refer to the 2020 Consolidated file (HC-224) and, for more detail, the documentation for the prescription drug file (HC-220A) when analyzing prescription drug spending over time.

As always, it is recommended that data users review relevant sections of the documentation for descriptions of these types of changes that might affect the interpretation of changes over time before undertaking trend analyses.

Analysts may wish to consider using techniques to smooth or stabilize analyses of trends using MEPS data such as comparing pooled time periods (e.g. 1996-1997 versus 2011-2012), working with moving averages, or using modeling techniques with several consecutive years of MEPS data to test the fit of specified patterns over time.

Finally, statistical significance tests should be conducted to assess the likelihood that observed trends are not attributable to sampling variation. In addition, researchers should be aware of the impact of multiple comparisons on Type I error. Without making appropriate allowance for multiple comparisons, undertaking numerous statistical significance tests of trends increases the likelihood of concluding that a change has taken place when one has not.

4.0 Strategies for Estimation

This file is constructed for estimation of utilization, expenditures, and sources of payment for Other Medical expenditures and to allow for estimates for the number of persons who obtained medical items in 2020.

4.1 Basic Estimates of Utilization, Expenditures, and Sources of Payment

This file contains round-specific expenditure data on purchases of each type of medical equipment, supplies, and services (see description below and OMTYPE_M18 variable in codebook for more details). Data are not collected on the actual number of purchases of the items/services represented on this file, so it is not possible to estimate the average expenditure per unit of service.

4.1.1 Type of Records on File (OMTYPE_M18)

Records for purchases of insulin and diabetic supplies were included in the Other Medical Expenses event files for 1996-2004. Beginning with the 2005 file, these records were excluded from the Other Medical Expenses event file since the expenditures have always been included on

the Prescribed Medicines file. The Prescribed Medicines file is a more appropriate source for estimates of both utilization and expenditures for insulin and diabetic supplies.

Each record on this file contains person-specific information on total expenditures during a specific round for a given category of medical equipment, services, and supplies (a maximum of 3 records per category of medical equipment for a sample person). Variables for annual expenditure data for each category of medical equipment, services, and supplies (obtained by cumulating across round-specific data in this file) are included on the annual Full Year Consolidated File.

Estimates of the total number of persons with expenditures for an item during the year are the sum of the weight variable (PERWT20F) across relevant records (e.g., for ambulance services, records where OMTYPE_M18 = 2). Estimates of expenditure variables must be weighted by PERWT20F to be nationally representative. For example, the estimate for the total expenditures for ambulance services paid out of pocket is produced by summing the product of the variables PERWT20F and OMSF20X across all the events in the file where OMTYPE_M18 = 2 as follows (the subscript 'j' identifies each event and represents a numbering of events from 1 through the total number of events in the file):

$\sum W_j X_j$, where

$W_j = \text{PERWT20F}_j$ (full-year weight for the person associated with event j) and

$X_j = \text{OMSF20X}_j$ (amount paid by self/family for event j) where OMTYPE_M18 = 2.

The estimate for the average expenditures for ambulance services paid out of pocket per person per round with that type of expense is produced as follows (the subscript 'j' identifies each event and represents a numbering of events from 1 through the total number of events in the file):

$(\sum W_j X_j) / (\sum W_j)$, where

$W_j = \text{PERWT20F}_j$ (full-year weight for the person associated with event j) and

$X_j = \text{OMSF20X}_j$ (amount paid by self/family for event j) where OMTYPE_M18 = 2.

This type of estimate and corresponding [standard error \(SE\)](#) can be derived using an appropriate computer software package for complex survey analysis such as SAS, Stata, SUDAAN, R, or SPSS. Variables are contained on the full-year annual file for aggregate expenditures across all of these types of services/items (OMTYPE_M18 = 1, 2, 3, or 4), but it is necessary to use this file to produce an annual estimate for a specific category of service. Small sample sizes make it advisable to pool multiple years of MEPS data to produce statistically reliable estimates for some of the items.

4.2 Variables with Missing Values

It is essential that the analyst examine all variables for the presence of negative values used to represent missing values. For continuous or discrete variables, where means or totals may be taken, it may be necessary to set negative values to values appropriate to the analytic needs. That is, the analyst should either impute a value or set the value to one that will be interpreted as missing by the software package used. For categorical and dichotomous variables, the analyst may want to consider whether to recode or impute a value for cases with negative values or whether to exclude or include such cases in the numerator and/or denominator when calculating proportions.

Methodologies used for the editing/imputation of expenditure variables (e.g., sources of payment and zero expenditures) are described in “Data Editing and Imputation Methodologies of Expenditure Variable.”

4.3 Variance Estimation (VARPSU, VARSTR)

The MEPS has a complex sample design. To obtain estimates of variability (such as the standard error of sample estimates or corresponding confidence intervals) for MEPS estimates, analysts need to take into account the complex sample design of MEPS for both person-level and family-level analyses. Several methodologies have been developed for estimating standard errors for surveys with a complex sample design, including the Taylor-series linearization method, balanced repeated replication, and jackknife replication. Various software packages provide analysts with the capability of implementing these methodologies. MEPS analysts most commonly use the Taylor Series approach. Although this data file does not contain replicate weights, the capability of employing replicate weights constructed using the Balanced Repeated Replication (BRR) methodology is also provided if needed to develop variances for more complex estimators (see Section 4.3.2).

4.3.1 Taylor-series Linearization Method

The variables needed to calculate appropriate standard errors based on the Taylor-series linearization method are included on this file as well as all other MEPS public use files. Software packages that permit the use of the Taylor-series linearization method include SUDAAN, R, Stata, SAS (version 8.2 and higher), and SPSS (version 12.0 and higher). For complete information on the capabilities of a package, analysts should refer to the corresponding software user documentation.

Using the Taylor-series linearization method, variance estimation strata and the variance estimation PSUs within these strata must be specified. The variables VARSTR and VARPSU on this MEPS data file serve to identify the sampling strata and primary sampling units required by the variance estimation programs. Specifying a “with replacement” design in one of the previously mentioned computer software packages will provide estimated standard errors appropriate for assessing the variability of MEPS survey estimates. It should be noted that the number of degrees of freedom associated with estimates of variability indicated by such a

package may not appropriately reflect the number available. For variables of interest distributed throughout the country (and thus the MEPS sample PSUs), one can generally expect to have at least 100 degrees of freedom associated with the estimated standard errors for national estimates based on this MEPS database.

Prior to 2002, MEPS variance strata and PSUs were developed independently from year to year, and the last two characters of the strata and PSU variable names denoted the year. However, beginning with the 2002 Point-in-Time PUF, the variance strata and PSUs were developed to be compatible with all future PUFs until the NHIS design changed. Thus, when pooling data across years 2002 through the Panel 11 component of the 2007 files, the variance strata and PSU variables provided can be used without modification for variance estimation purposes for estimates covering multiple years of data. There were 203 variance estimation strata, each stratum with either two or three variance estimation PSUs.

From Panel 12 of the 2007 files, a new set of variance strata and PSUs were developed because of the introduction of a new NHIS design. There are 165 variance strata with either two or three variance estimation PSUs per stratum, starting from Panel 12. Therefore, there are a total of 368 (203+165) variance strata in the 2007 Full-Year file as it consists of two panels that were selected under two independent NHIS sample designs. Since both MEPS panels in the full-year files from 2008 through 2016 are based on the next NHIS design, there are only 165 variance strata. These variance strata (VARSTR values) have been numbered from 1001 to 1165 so that they can be readily distinguished from those developed under the former NHIS sample design in the event that data are pooled for several years.

As discussed, a complete change was made to the NHIS sample design in 2016, effectively changing the MEPS design beginning with calendar year 2017. There were 117 variance strata originally formed under this new design intended for use until the next fully new NHIS design was implemented. In order to make the pooling of data across multiple years of MEPS more straightforward, the numbering system for the variance strata has changed. Those strata associated with the new design (implemented in 2016) were numbered from 2001 to 2117.

However, the new NHIS sample design implemented in 2016, was further modified in 2018. With the modification in the 2018 NHIS sample design, the MEPS variance structure for the 2019 Full Year file has also had to be modified, reducing the number of variance strata to 105. Consistency was maintained with the prior structure in that the 2019 Full Year file variance strata were also numbered within the range of values from 2001-2117, although there are now gaps in the values assigned within this range.

Some analysts may be interested in pooling data across multiple years of MEPS data. As noted on the cover page of this document, due to data quality issues arising from collecting data during the COVID-19 pandemic in 2020, caution should be taken when interpreting the results of such pooling.

If pooling is to be undertaken, it should be noted that, to obtain appropriate standard errors when doing so, it is necessary to specify a common variance structure. Prior to 2002, each annual MEPS public use file was released with a variance structure unique to the particular MEPS sample in that year. Starting in 2002, the annual MEPS public use files were released with a

common variance structure that allowed users to pool data from 2002 through 2018. However, with the need to modify the variance structure beginning with 2019, this can no longer be routinely done.

To ensure that variance strata are identified appropriately for variance estimation purposes when pooling MEPS data across several years, one can proceed as follows:

1. When pooling any year from 2002 through 2018, use the variance strata numbering as is.
2. When pooling (a) any year from 1996 to 2001 with any year from 2002 or later, or (b) 2019 and beyond with any earlier year, use the pooled linkage public use file HC-036 that contains the proper variance structure. The HC-036 file is updated every year so that appropriate variance structures are available with pooled data. Further details on the HC-036 file can be found in the public use documentation of the HC-036 file.

4.3.2 Balanced Repeated Replication (BRR) Method

BRR replicate weights are not provided on this MEPS PUF for the purposes of variance estimation. However, a file containing a BRR replication structure is made available so users can form replicate weights, if desired, from the final MEPS weight to compute variances of MEPS estimates using either BRR or Fay's modified BRR (Fay 1989) methods. The replicate weights are useful to compute variances of complex non-linear estimators for which a Taylor linear form is not easy to derive and not available in commonly used software. For instance, it is not possible to calculate the variances of a median or the ratio of two medians using the Taylor linearization method. For these types of estimators, users may calculate a variance using BRR or Fay's modified BRR methods. However, it should be noted that the replicate weights have been derived from the final weight through a shortcut approach. Specifically, the replicate weights are not computed starting with the base weight and all adjustments made in different stages of weighting are not applied independently in each replicate. Thus, the variances computed using this one-step BRR do not capture the effects of all weighting adjustments that would be captured in a set of fully developed BRR replicate weights. The Taylor Series approach does not fully capture the effects of the different weighting adjustments either.

The dataset, HC-036BRR, MEPS 1996-2018 Replicates for Variance Estimation File, contains the information necessary to construct the BRR replicates. It contains a set of 128 flags (BRR1-BRR128) in the form of half sample indicators, each of which is coded 0 or 1 to indicate whether the person should or should not be included in that particular replicate. These flags can be used in conjunction with the full-year weight to construct the BRR replicate weights. For analysis of MEPS data pooled across years, the BRR replicates can be formed in the same way using the HC-036, MEPS 1996-2018 Pooled Linkage Variance Estimation File. For more information about creating BRR replicates, users can refer to the documentation for the [HC-036BRR pooled linkage file](#) on the AHRQ website.

5.0 Merging/Linking MEPS Data Files

Data from this file can be used alone or in conjunction with other files for different analytic purposes. This section summarizes various scenarios for merging/linking MEPS event files. The set of households selected for MEPS is a subsample of those participating in the National Health Interview Survey (NHIS), thus, each MEPS panel can also be linked back to the previous year's NHIS public use data files. For information on obtaining MEPS/NHIS link files please see the [MEPS website](#).

5.1 Linking to the Person-Level File

Merging characteristics of interest from other MEPS files (e.g., 2020 Full Year Consolidated File or 2020 Prescribed Medicines File) expands the scope of potential estimates. For example, to estimate the expenditures for medical equipment, visual aids, etc. for persons with specific demographic characteristics (such as age, race, and sex), population characteristics from a person-level file need to be merged onto the Other Medical Expenses event file. This procedure is shown below.

1. Create data set PERSX by sorting the 2020 Full Year Consolidated File, by the person identifier, DUPERSID. Keep only variables to be merged onto the other medical events file and DUPERSID.
2. Create data set OMEXP by sorting the other medical event file by person identifier, DUPERSID.
3. Create final data set NEWOME by merging these two files by DUPERSID, keeping only records on the other medical event file.

The following is an example of SAS code that completes these steps:

```
PROC SORT DATA=HCXXX (KEEP=DUPERSID AGE31X AGE42X AGE53X
SEX RACEV1X EDUCYR HIDEG) OUT=PERSX;
    BY DUPERSID;
RUN;
PROC SORT DATA=OMEXP;
    BY DUPERSID;
RUN;
DATA NEWOME;
MERGE OMEXP (IN=A) PERSX (IN=B);
    BY DUPERSID;
IF A;
RUN;
```


5.2 Linking to the Prescribed Medicines File

The RXLK file provides a link from the MEPS event files to the 2020 Prescribed Medicines Event File. When using RXLK, data users/analysts should keep in mind that one other medical record can link to more than one prescribed medicine record. Conversely, a prescribed medicine event may link to more than one other medical record. When this occurs, it is up to the data user/analyst to determine how the prescribed medicine expenditures should be allocated among those medical events.

5.3 Linking to the Medical Conditions File

Conditions data are not collected for Other Medical events; therefore, this file cannot be linked to the Conditions File.

References

- Bramlett, M.D., Dahlhamer, J.M., & Bose, J. (2021, September). [*Weighting Procedures and Bias Assessment for the 2020 National Health Interview Survey*](#). Centers for Disease Control and Prevention.
- Chowdhury, S.R., Machlin, S.R., & Gwet, K.L. [*Sample Designs of the Medical Expenditure Panel Survey Household Component, 1996-2006 and 2007-2016. Methodology Report #33.*](#) January 2019. Agency for Healthcare Research and Quality, Rockville, MD.
- Cohen, S.B. (1996). The Redesign of the Medical Expenditure Panel Survey: A Component of the DHHS Survey Integration Plan. *Proceedings of the COPAFS Seminar on Statistical Methodology in the Public Service*.
- [*Current Population Survey: 2021 Annual Social and Economic \(ASEC\) Supplement*](#). (2021). U.S. Census Bureau.
- Dahlhamer, J.M., Bramlett, M.D., Maitland, A., & Blumberg, S.J. (2021). [*Preliminary evaluation of nonresponse bias due to the COVID-19 pandemic on National Health Interview Survey estimates, April-June 2020*](#). National Center for Health Statistics.
- Daily, D., Cantwell, P.J., Battle, K., & Waddington, D.G. (2021, October 27), [*An Assessment of the COVID-19 Pandemic's Impact on the 2020 ACS 1-Year Data*](#). U.S. Census Bureau.
- Fay, R.E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. *Proceedings of the Survey Research Methods Sections, ASA*, 212-217.
- Lau, D.T., Sosa, P., Dasgupta, N., & He, H. (2021). [*Impact of the COVID-19 Pandemic on Public Health Surveillance and Survey Data Collections in the United States*](#). *American Journal of Public Health*, 111 (12), pp. 2118-2121.
- Monheit, A.C., Wilson, R., and Arnett, III, R.H. (Editors) (1999). *Informing American Health Care Policy*. Jossey-Bass Inc., San Francisco.
- Rothbaum, J. & Bee, A. (2020). Coronavirus Infects Surveys, Too: Nonresponse Bias During the Pandemic in the CPS ASEC (SEHSD Working Paper Number 2020-10). U.S. Census Bureau.
- Rothbaum, J. & Bee, A. (2021, May 3). [*Coronavirus Infects Surveys, Too: Survey Nonresponse Bias and the Coronavirus Pandemic*](#). U.S. Census Bureau.
- Rothbaum, J., Eggleston, J., Bee, A., Klee, M., & Mendez-Smith, B. (2021). [*Addressing Nonresponse Bias in the American Community Survey During the Pandemic Using Administrative Data*](#). U.S. Census Bureau.
- Shah, B.V., Barnwell, B.G., Bieler, G.S., Boyle, K.E., Folsom, R.E., Lavange, L., Wheelless, S.C., and Williams, R. (1996). *Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0*. Research Triangle Park, NC: Research Triangle Institute.

Villa Ross, C.A., Shin, H.B., & Marlay, M.C. (2021, October 27). [Pandemic Impact on 2020 American Community Survey 1-Year Data](#). U.S. Census Bureau.

Zuvekas, S.H. & Kashihara, D. (2021). [The Impacts of the COVID-19 Pandemic on the Medical Expenditure Panel Survey](#). *American Journal of Public Health*, 111 (12), pp. 2157-2166.

D. Variable-Source Crosswalk

FOR MEPS HC-220C: 2020 OTHER MEDICAL EXPENSES

Survey Administration Variables

Variable	Description	Source
DUID	Panel # + Encrypted DU identifier	Assigned in sampling
PID	Person number	Assigned in sampling
DUPERSID	Person ID (DUID + PID)	Assigned in sampling
EVNTIDX	Event ID	Assigned in Sampling
EVENTRN	Event round number	CAPI derived
PANEL	Panel number	Constructed

Other Medical Events Variables

Variable	Description	Source
OMTYPE_M18	Other medical expense type	OM10, 30, 40, 50

Imputed Expenditure Variables

Variable	Description	Source
OMSF20X	Amount paid, family (Imputed)	CP Section (Edited)
OMMR20X	Amount paid, Medicare (Imputed)	CP Section (Edited)
OMMD20X	Amount paid, Medicaid (Imputed)	CP Section (Edited)
OMPV20X	Amount paid, private insurance (Imputed)	CP Section (Edited)
OMVA20X	Amount paid, Veterans/CHAMPVA (Imputed)	CP Section (Edited)
OMTR20X	Amount paid, TRICARE (Imputed)	CP Section (Edited)
OMOF20X	Amount paid, other federal (Imputed)	CP Section (Edited)
OMSL20X	Amount paid, state & local government (Imputed)	CP Section (Edited)
OMWC20X	Amount paid, workers' compensation (Imputed)	CP Section (Edited)
OMOT20X	Amount paid, other insurance (Imputed)	CP Section (Edited)
OMXP20X	Sum of OMSF20X-OMOT20X (Imputed)	Constructed

Variable	Description	Source
OMTC20X	Household reported total charge (Imputed)	CP Section (Edited)
IMPFLAG	Imputation status	Constructed

Weights Variables

Variable	Description	Source
PERWT20F	Expenditure file person weight, 2020	Constructed
VARSTR	Variance estimation stratum, 2020	Constructed
VARPSU	Variance estimation PSU, 2020	Constructed