MEPS HC-213G:
2019 Office-Based Medical Provider Visits

June 2021

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

Section					Page
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-2
		2.1	Codebook Structure		C-3
		2.2	Reserved Codes		C-4
		2.3	Codebook Format		C-4
		2.4	Variable Source and Naming Conventions		C-5
			2.4.1	General	C-5
			2.4.2	Expenditure and Source of Payment Variables	C-5
		2.5	File Contents		C-6
			2.5.1	Survey Administration Variables	C-6
				2.5.1.1 Person Identifiers (DUID, PID, DUPERSID)	C-6
				2.5.1.2 Record Identifiers (EVNTIDX, FFEEIDX)	C-7
				2.5.1.3 Round Indicator (EVENTRN)	C-7
				2.5.1.4 Panel Indicator (PANEL)	C-7
			2.5.2	MPC Data Indicator (MPCELIG, MPCDATA)	C-7
			2.5.3	Office-Based Medical Provider Visit Variables	C-7
				2.5.3.1 Date of Visit (OBDATEMM - OBDATEYR)	C-8

Section			Page
	2.5.3.2	Visit Details (SEEDOC_M18-VSTRELCN_M18)	C-8
	2.5.3.3	Procedures, Services, and Prescription Medicines (LABTEST_M18-MEDPRESC)	C-8
2.5.4		Clinical Classification Software Refined	C-8
2.5.5		Flat Fee Variables (FFEEIDX, FFOBTYPE, FFBEF19, FFTOT20)	C-9
	2.5.5.1	Definition of Flat Fee Payments	C-9
	2.5.5.2	Flat Fee Variable Descriptions	C-9
	2.5.5.3	Caveats of Flat Fee Groups	C-10
2.5.6		Expenditure Data	C-10
	2.5.6.1	Definition of Expenditures	C-10
	2.5.6.2	Data Editing and Imputation Methodologies of Expenditure Variables	C-11
	2.5.6.3	Capitation Imputation	C-13
	2.5.6.4	Imputation Flag (IMPFLAG)	C-13
	2.5.6.5	Flat Fee Expenditures	C-13
	2.5.6.6	Zero Expenditures	C-14
	2.5.6.7	Discount Adjustment Factor	C-14
	2.5.6.8	Sources of Payment	C-14
	2.5.6.9	Office-Based Expenditure Variables (OBSF19X - OBTC19X)	C-15
2.5.7		Rounding	C-15

Section			Page
3.0	Sample Weight (PERWT19F)		C-15
	3.1	Overview	C-15
	3.2	Details on Person Weight Construction	C-16
		3.2.1 MEPS Panel 23 Weight Development Process	C-16
		3.2.2 MEPS Panel 24 Weight Development Process	C-17
		3.2.3 The Final Weight for 2019	C-17
		3.2.4 Coverage	C-18
	3.3	Using MEPS Data for Trend Analysis	C-18
4.0	Strategies for Estimation		C-19
	4.1	Developing Event-Level Estimates	C-19
	4.2	Person-Based Estimates for Office-Based Visits	C-21
	4.3	Variables with Missing Values	C-21
	4.4	Variance Estimation (VARSTR, VARPSU)	C-22
		4.4.1 Taylor-series Linearization Method	C-22
		4.4.2 Balanced Repeated Replication (BRR) Method	C-23
5.0	Merging/Linking MEPS Data Files		C-24
	5.1	Linking to the Person-Level File	C-24
	5.2	Linking to the Prescribed Medicines File	C-25
	5.3	Linking to the Medical Conditions File	C-25
References			C-26
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, 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. The linkage of the MEPS to the previous year's NHIS provides additional data for longitudinal analytic purposes.

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.

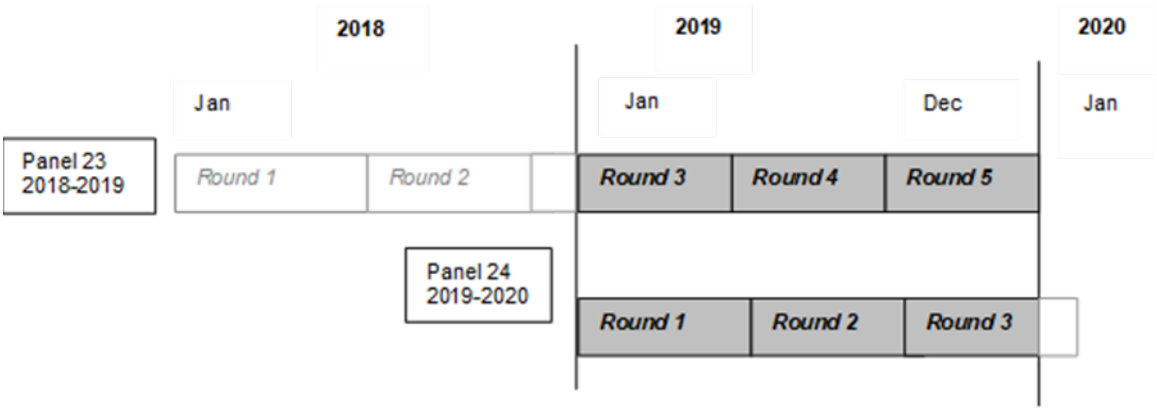
As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports, micro data files, and tables via the MEPS website.

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 one in a series of public use event files from the 2019 Medical Expenditure Panel Survey (MEPS) Household (HC) and Medical Provider Components (MPC). Released as an ASCII data file (with related SAS, Stata, SPSS, and R programming statements), SAS data set, SAS transport file, Stata data set, and Excel file the 2019 Office-Based Medical Provider Visits public use event file provides detailed information on office-based provider visits for a nationally representative sample of the civilian noninstitutionalized population of the United States. Data from the office-based provider events file can be used to make estimates of office-based provider utilization and expenditures for calendar year 2019. The file contains 45 variables and has a logical record length of 213 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 from the 2019 portion of Round 3 and Rounds 4 and 5 for Panel 23, as well as Rounds 1, 2 and the 2019 portion of Round 3 for Panel 24 (i.e., the rounds for the MEPS panels covering calendar year 2019).



Each record on this event file represents a unique office-based provider event; that is, an office-based provider event reported by the household respondent. Office-based events reported in Panel 24 Round 3 and known to have occurred after December 31, 2019 are not included on this file. Utilization counts of office-based provider visits are based entirely on household reports. Information from the MPC is used to supplement expenditure payment data on the office-based provider file, reported by the household, and does not affect use estimates.

Data from this event file can be merged with other 2019 MEPS HC data files for purposes of appending person-level data such as demographic characteristics or health insurance coverage to each office-based provider visit record on the current file.

This file can also be used to construct summary variables of expenditures, sources of payment, and related aspects of office-based provider visits for calendar year 2019. Aggregate annual

person-level information on the use of office-based providers and other health services is provided on the MEPS 2019 Full Year Consolidated Data File, where each record represents a MEPS sampled person.

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

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

Any variables not found on this file but released on previous years' files may have been excluded because they contained only missing data.

For more information on the MEPS HC sample design, see Chowdhury et al (2019). A copy of the MEPS HC survey instruments used to collect the information on the office-based provider file is available on the MEPS website.

2.0 Data File Information

The 2019 Office-Based Medical Provider public use data set consists of one event-level data file. The file contains characteristics associated with the office-based (OB) event and imputed expenditure data.

The Office-Based Provider public use data set contains 188,782 office-based provider event records; of these records, 186,151 are associated with persons having a positive person-level weight (PERWT19F). This file includes office-based provider event records for all household members who resided in eligible responding households and for whom at least one office-based provider event was reported.

Each record represents one household-reported office-based provider event that occurred during calendar year 2019. Office-based provider visits known to have occurred after December 31, 2019 are not included on this file. Some household members may have multiple events and thus will be represented in multiple records on this file. Other household members may have had no events reported and thus will have no records on this file. These data were collected during the 2019 portion of Round 3, and Rounds 4 and 5 for Panel 23, as well as Rounds 1, 2, and the 2019

portion of Round 3 for Panel 24 of the MEPS Household Component. The persons represented on this file had to meet either (a) or (b):

a) Be classified as a key in-scope person who responded for his or her entire period of 2019 eligibility (i.e., persons with a positive 2019 full-year person-level weight (PERWT19F > 0)), or

b) Be an eligible member of a family all of whose key in-scope members have a positive person-level weight (PERWT19F > 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 (FAMWT19F >0). Note that FAMIDYR and FAMWT19F are variables on the 2019 Full Year Consolidated Data File.

Persons with no office-based medical provider visit events for 2019 are not included on this event-level OB file but are represented on the person-level 2019 Full Year Population Characteristics file.

Each office-based medical provider visit event record includes the following: date of the event; type of provider seen; type of care received; type of services (i.e., lab test, sonogram or ultrasound, x-rays, etc.) received, medicines prescribed during the event; flat fee information; imputed sources of payment; total payment and total charge of the office-based event expenditure; a full-year person-level weight; variance strata; and variance PSU.

To append person-level information such as demographic or health insurance coverage to each event record, data from this file can be merged with 2019 MEPS HC person-level data (e.g. Full Year Consolidated or Full Year Population Characteristics file) using the person identifier, DUPERSID. The office-based medical provider visit events can also be linked to the MEPS 2019 Medical Conditions File and MEPS 2019 Prescribed Medicines File. Please see Section 5.0 for details on how to merge MEPS data files.

2.1 Codebook Structure

For most variables on the office-based provider 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 identifiers
- Unique office-based medical provider visit event identifiers
- Office-based medical provider visit characteristic variables
- 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 office-based medical provider visits codebook describes an ASCII data set (although the data are also being provided in an Excel file, Stata data set, SAS data set, and SAS transport 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".

Beginning in 2018, 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 General

Variables contained on this file were derived from the HC survey questionnaire itself, derived from the MPC data collection instrument, derived from CAPI, or assigned in sampling. 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 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:

   - MV – Office-Based Medical Provider Visits section
   - FF – Flat Fee section
   - CP – Charge Payment section

3. Variables constructed from multiple questions using complex algorithms are labeled "Constructed" in the "Source" column; and

4. Variables which 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, 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 the 10 sources of payment are named in the following way:

The first two characters indicate the type of event:

IP - inpatient stay ER - emergency room visit

HH - home health visit

OM - other medical equipment

OB - office-based visit

OP - outpatient visit

DV - dental visit

RX - prescribed medicine

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

SF - self or family

MR - Medicare

MD - Medicaid

PV - private insurance

VA - Veterans Administration/CHAMPVA

TR - TRICARE

OF - other federal government

SL - state/local government

WC - Workers' Compensation

OT - other insurance

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 (19). The seventh character, "X", indicates whether the variable is edited/imputed.

For example, OBSF19X is the edited/imputed amount paid by self or family for an office-based medical provider visit expenditure incurred in 2019.

Beginning in 2019, the expenditure variables OR (other private) and OU (other public) are dropped from the file.

2.5 File Contents

2.5.1 Survey Administration Variables

2.5.1.1 Person Identifiers (DUID, PID, DUPERSID)

The dwelling unit ID (DUID) is a seven-digit ID number consisting of a 2-digit panel number followed by a five-digit random number assigned after the case was sampled for MEPS. The three-digit person number (PID) uniquely identifies each person within the dwelling unit. The ten-character variable DUPERSID uniquely identifies each person represented on the file and is the combination of the variables DUID and PID. Beginning in 2018, the lengths of the ID variables begin with the 2-digit panel number.

For detailed information on dwelling units and families, please refer to the documentation for the 2019 Full Year Population Characteristics.

2.5.1.2 Record Identifiers (EVNTIDX, FFEEIDX)

EVNTIDX uniquely identifies each office-based medical provider visit event (i.e., each record on the office-based medical provider visits file) and is the variable required for linking office-based medical provider visit events to data files containing details on conditions and/or prescribed medicines (MEPS 2019 Medical Condition File and MEPS 2019 Prescribed Medicines File, respectively). Beginning in 2018, 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 2019 Appendix File, HC-213I.

FFEEIDX is a constructed variable that uniquely identifies a flat fee group, that is, all events that were part of a flat fee payment. For example, pregnancy is typically covered in a flat fee arrangement where the prenatal visits, the delivery, and the postpartum visits are all covered under one flat fee dollar amount. These events (the prenatal visit, the delivery, and the postpartum visits) would have the same value for FFEEIDX. FFEEIDX identifies a flat fee payment that was identified using information from the Household Component. A "mixed" flat fee group could contain both outpatient and office-based visits. Only outpatient and office-based events are allowed in a mixed bundle. Please note that FFEEIDX should be used to link up the outpatient and office-based events in order to determine the full set of events that are part of a flat fee group.

2.5.1.3 Round Indicator (EVENTRN)

EVENTRN indicates the round in which the office-based event was reported. Please note that Rounds 3, 4, and 5 are associated with MEPS survey data collected from Panel 23. Likewise, Rounds 1, 2, and 3 are associated with data collected from Panel 24.

2.5.1.4 Panel Indicator (PANEL)

PANEL is a constructed variable used to specify the panel number for the person. PANEL will indicate either Panel 23 or Panel 24 for each person on the file. Panel 23 is the panel that started in 2018, and Panel 24 is the panel that started in 2019.

2.5.2 MPC Data Indicator (MPCELIG, MPCDATA)

MPCELIG is a constructed variable that indicates whether the office-based provider visit was eligible for MPC data collection. MPCDATA is a constructed variable that indicates whether or not MPC data were collected for the office-based provider.

2.5.3 Office-Based Medical Provider Visit Variables

The file contains variables describing office-based medical provider visit events reported by respondents in the Medical Provider Visits section of the MEPS HC survey questionnaire.

2.5.3.1 Date of Visit (OBDATEMM - OBDATEYR)

There are two variables that, together, indicate the month and year an office-based provider visit occurred (OBDATEMM and OBDATEYR, respectively). These variables have not been edited or imputed.

2.5.3.2 Visit Details (SEEDOC_M18- VSTRELCN_M18)

The questionnaire establishes whether the person saw or spoke to a medical doctor (SEEDOC_M18). If the person talked to a medical doctor, the respondent is asked to specify the type (DRSPLTY_M18) and other health professional type (MEDPTYPE_M18) is set to -1, "INAPPLICABLE". If during the medical visit the patient did not see a specialty doctor (DRSPLTY_M18) or, if the person did not see a physician (i.e., a medical doctor), the respondent was asked to identify the type of medical person seen (MEDPTYPE_M18). Whether or not any medical doctors worked at the visit location (DOCATLOC), the type of care the person received (VSTCTGRY), and whether or not the visit or telephone call was related to a specific condition (VSTRELCN_M18) were also determined. Note that response categories with small frequencies may have been recoded to other categories for confidentiality reasons.

2.5.3.3 Procedures, Services, and Prescription Medicines (LABTEST_M18-MEDPRESC)

Services received during the visit included whether or not the person received lab tests (LABTEST_M18), a sonogram or ultrasound (SONOGRAM_M18), x-rays (XRAYS_M18), a mammogram (MAMMOG_M18), an MRI or a CAT scan (MRI_M18), an electrocardiogram/an electroencephalogram (EKG_M18), or a vaccination (RCVVAC_M18). Minimal editing was done across treatment, services, and procedures to ensure consistency across "inapplicable," "don't know," "refused," and "no services received" values.

Whether or not a surgical procedure was performed during the visit was asked (SURGPROC).

Finally, the questionnaire determined if a medicine was prescribed for the person during the visit (MEDPRESC). For a repeat visit event group, if a prescribed medicine is linked to the stem event (MEDPRESC=1), the value of MEDPRESC is copied to the leaf events without linking the leaf events to the prescribed medicine, then MEDPRESC=1 was recoded to -15 for all leaf events.

2.5.4 Clinical Classification Software Refined

Information on household-reported medical conditions (ICD-10-CM condition codes) and aggregated clinically meaningful categories generated using Clinical Classification Software Refined (CCSR) for each office-based medical provider visit are not provided on this file. For information on the ICD-10-CM condition codes and associated CCSR codes, see the MEPS 2019 Medical Conditions File.

2.5.5 Flat Fee Variables (FFEEIDX, FFOBTYPE, FFBEF19, FFTOT20)

2.5.5.1 Definition of Flat Fee Payments

A flat fee is the fixed dollar amount a person is charged for a package of services provided during a defined period of time. An example would be an obstetrician's fee covering a normal delivery, and the associated pre- and post-natal care. A flat fee group is the set of medical services (i.e., events) that are covered under the same flat fee payment. The flat fee groups represented on the office-based provider file include flat fee groups where at least one of the health care events, as reported by the HC respondent, occurred during 2019. By definition, a flat fee group can span multiple years and/or event types (only outpatient department visits and physician office visits). Furthermore, a single person can have multiple flat fee groups.

2.5.5.2 Flat Fee Variable Descriptions

Flat Fee ID (FFEEIDX)

As noted earlier in Section 2.5.1.2 "Record Identifiers," the variable FFEEIDX uniquely identifies all events that are part of the same flat fee group for a person. On any 2019 MEPS event file, every event that was a part of a specific flat fee group will have the same value for FFEEIDX. Note that prescribed medicine and home health events are never included in a flat fee group and FFEEIDX is not a variable on those event files.

Flat Fee Type (FFOBTYPE)

FFOBTYPE indicates whether the 2019 office-based medical provider visit event is the "stem" or "leaf" of a flat fee group. A stem (records with FFOBTYPE = 1) is the initial medical service (event) which is followed by other medical events that are covered under the same flat fee payment. The leaves of the flat fee group (records with FFOBTYPE = 2) are those medical events that are tied back to the initial medical event (the stem) in the flat fee group. These "leaf" records have their expenditure variables set to zero. For the office-based visits that are not part of a flat fee payment, the FFOBTYPE is set to -1, "INAPPLICABLE."

Counts of Flat Fee Events that Cross Years (FFBEF19, FFTOT20)

As described in Section 2.5.5.1, a flat fee payment covers multiple events and the multiple events could span multiple years. For situations where the office-based medical provider visit occurred in 2018 as a part of a group of events, and some of the events occurred before 2019, counts of the known events are provided on the office-based medical provider visit event file record. Variables that indicate events occurring before or after 2019 are as follows:

FFBEF19 - total number of pre-2019 events in the same flat fee group as the 2019 office-based medical provider visit. This count would not include the 2019 office-based medical visit(s).

FFTOT20 - the number of 2020 office-based events expected to be in the same flat fee group as the office-based medical provider visit event(s) that occurred in 2019.

If there are no 2018 events on the file, FFBEF19 will be omitted. Likewise, if there are no 2020 events on the file, FFTOT20 will be omitted. If there are no flat fee data related to the records in this file, FFEEIDX and FFOBTYPE will be omitted as well. Please note that the crosswalk in this document lists all possible flat fee variables.

2.5.5.3 Caveats of Flat Fee Groups

Data users/analysts should note that flat fee payments are common on the office-based medical provider visits file. There are 2,175 office-based medical provider visit events that are identified as being part of a flat fee payment group. In order to correctly identify all events that are part of a flat fee group, the user should link all MEPS events, except those in the prescribed medicine file, using the variable FFEEIDX. In general, every flat fee group should have an initial visit (stem) and at least one subsequent visit (leaf). There are some situations where this is not true. For some of these flat fee groups, the initial visit reported occurred in 2019, but the remaining visits that were part of this flat fee group occurred in 2020. In this case, the 2019 flat fee group represented on this file would consist of one event (the stem). The 2020 leaf events that are part of this flat fee group are not represented on this file. Similarly, the household respondent may have reported a flat fee group where the initial visit began in 2018 but subsequent visits occurred during 2019. In this case, the initial visit would not be represented on the file. This 2019 flat fee group would then consist only of one or more leaf records and no stem. Another reason for which a flat fee group would not have a stem and at least one leaf record is that the stem or leaves could have been reported as different event types. Outpatient and office-based medical provider visits are the only two event types allowed in a single flat fee group. The stem may have been reported as an outpatient department visit and the leaves may have been reported as office-based medical provider visits. Please note that the crosswalk in this document lists all possible flat fee variables.

2.5.6 Expenditure Data

2.5.6.1 Definition of Expenditures

Expenditures on this file refer to what is paid for health care services. More specifically, expenditures in MEPS are defined as the sum of payments for care received, 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, the estimates do not incorporate any payment not directly tied to specific medical care visits, such as bonuses or retrospective payment adjustments paid by third party payers. Currently, 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 is it directly comparable to the resource costs of those services or the expenditures defined in the 1987 NMES (for details on expenditure definitions, see 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, and also are available in Zuvekas and Cohen, 2002. For more information, see the Data Center section of the MEPS website. If examining trends in MEPS expenditures, please refer to Section 3.3 for more information.

2.5.6.2 Data Editing and Imputation Methodologies of Expenditure Variables

The expenditure data included on this file were derived from both the MEPS household (HC) and medical provider (MPC) components. The MPC contacted medical providers identified by household respondents. The charge and payment data from medical providers were used in the expenditure imputation process to supplement missing household data. For all office-based medical provider visits, MPC data were used if available; otherwise HC data were used. Missing data for office-based medical provider visits where HC data were not complete and MPC data were not collected, or MPC data were not complete, were derived through the imputation process.

General Data Editing Methodology

Logical edits were used to resolve internal inconsistencies and other problems in the HC and MPC 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, co-payments or charges reported as total payments, and reimbursed amounts that were reported as out-of-pocket 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 missing payment was used to impute the missing payment value. Within each event type file, separate imputations were performed for flat fee and simple events. Separate imputations were performed for visits to physicians (where MPCELIG=1) and visits to non-physician providers (where MPCELIG=2). After the imputations were finished, visits to physician and non-physician providers were combined into a single medical provider file.

The 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.

Office-Based Provider Visit Data Editing and Imputation

Expenditures for office-based provider visits were developed in a sequence of logical edits and imputations. "Household" edits were applied to sources and amounts of payment for all events reported by HC respondents. "MPC" edits were applied to provider-reported sources and amounts of payment for records matched to household-reported events. Both sets of edits were used to correct obvious errors (as described above) in the reporting of expenditures. After the data from each source were edited, a decision was made as to whether household- or MPC-reported information would be used in the final editing and in the predictive mean matching imputations for missing expenditures. The general rule was that MPC data would be used for events where a household-reported event corresponded to an MPC-reported event (i.e., a matched event), since providers usually have more complete and accurate data on sources and amounts of payment than households.

One of the more important edits separated flat fee events from simple events. This edit was necessary because groups of events covered by a flat fee (i.e., a flat fee bundle) were edited and imputed separately from individual events covered by a single charge (i.e., simple events). (See Section 2.5.5 for more details on flat fee groups).

Logical edits also were used to sort each event into a specific category for the imputations. Events with complete expenditures were flagged as potential donors 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 eight recipient categories in which all events had a common extent of missing data. However, for predictive mean matching imputations, the recipients were grouped into four categories based on the known status of total charge and the sources of payment: 1) Known charge but unknown payment status of at least one potential paying source, 2) Unknown charge and unknown payment status of at least one potential paying source, 3) Known charge and known status of all payment sources, and 4) Unknown charge and known status of all payment sources. Separate imputations were performed on events in each recipient group. For office-based events, the donor pool was restricted to events with complete expenditures from the MPC. To improve the reliability of imputation, current year donors and inflation-adjusted prior year donors are used for the predictive mean matching imputations.

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) would not be represented among incomplete events (recipients).

For office-based and outpatient events, the donor pool also included events originally reported by providers as paid on a capitated basis. To obtain the fee-for-service (FFS) equivalent payments for these capitated events, a "capitation imputation" was implemented (see the next section). Once imputed with the FFS equivalent payments, these events became donors for all other incomplete events, particularly for events reported by the household as services covered under managed care plans.

2.5.6.3 Capitation Imputation

The weighted sequential hot-deck procedure was used to estimate expenditures at the event-level for events that were paid on a per month per person (capitated) basis. The capitation imputation procedure was designed as a reasonable approach to complete event-level expenditures for persons in non-fee for service managed care plans. HMO events reported in the MPC as covered by capitation arrangements were imputed using similar completed HMO events paid on a fee-for-service, with total charge as a key variable. Then this fully completed set of MPC events was used in the donor pool for the main imputation process for cases in HMOs. By using this strategy, capitated HMO events were imputed as if the provider were reimbursed from the HMO on a discounted fee-for-service basis.

2.5.6.4 Imputation Flag (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 OB 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 and flat fee leaf events)

IMPFLAG = 1 complete HC data

IMPFLAG = 2 complete MPC data

IMPFLAG = 3 fully imputed

IMPFLAG = 4 partially imputed

IMPFLAG = 5 complete MPC data through capitation imputation

2.5.6.5 Flat Fee Expenditures

The approach used to count expenditures for flat fees was to place the expenditure on the first visit of the flat fee group. The remaining visits have zero payments. Thus, if the first visit in the flat fee group occurred prior to 2019, all of the events that occurred in 2019 will have zero payments. Conversely, if the first event in the flat fee group occurred at the end of 2019, the total

expenditure for the entire flat fee group will be on that event, regardless of the number of events it covered after 2019. See Section 2.5.5 for details on the flat fee variables.

2.5.6.6 Zero Expenditures

There are some medical events reported by respondents where the payments were zero. Zero payment events can occur in MEPS for the following reasons: (1) the visit was covered under a flat fee arrangement (flat fee payments are included only on the first event covered by the arrangement), (2) there was no charge for a follow-up visit, (3) the provider was never paid directly for services provided by an individual, insurance plan, or other source, (4) the charges were included in another bill, or (5) the event was paid through government or privately-funded research or clinical trials.

2.5.6.7 Discount Adjustment Factor

An adjustment was also applied to some HC-reported expenditure data because an evaluation of matched HC/MPC data showed that respondents who reported that charges and payments were equal were often unaware that insurance payments for the care had been based on a discounted charge. To compensate for this systematic reporting error, a weighted sequential hot-deck imputation procedure was implemented to determine an adjustment factor for HC-reported insurance payments when charges and payments were reported to be equal. As for the other imputations, selected predictor variables were used to form groups of donor and recipient events for the imputation process.

2.5.6.8 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.

Prior to 2019, for cases where reported insurance coverage and sources of payment are inconsistent, the positive amount from a source inconsistent with reported insurance coverage was moved to one or both of the source categories Other Private and Other Public. Beginning in 2019, this step is removed and the inconsistency between the payment sources and insurance coverage is intact - the amounts are not moved to Other Private and Other Public categories any more. The two source of payment categories, Other Private and Other Public, are no longer available.

2.5.6.9 Office-Based Expenditure Variables (OBSF19X - OBTC19X)

OBSF19X - OBOT19X are the 10 sources of payment. The 10 sources of payment are: self/family (OBSF19X), Medicare (OBMR19X), Medicaid (OBMD19X), private insurance (OBPV19X), Veterans /CHAMPVA (OBVA19X), TRICARE (OBTR19X), other federal sources (OBOF19X), state and local (non-federal) government sources (OBSL19X), Workers' Compensation (OBWC19X), and other insurance (OBOT19X). OBXP19X is the sum of the 10 sources of payment for the office-based expenditures, and OBTC19X is the total charge.

2.5.7 Rounding

Expenditure variables have been rounded to the nearest penny. Person-level expenditure information released on the MEPS 2019 Person-Level Use and Expenditure File will be rounded to the nearest dollar. It should be noted that using the MEPS 2019 event files to create person-level totals will yield slightly different totals than those found on the person-level expenditure 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 Sample Weight (PERWT19F)

3.1 Overview

There is a single full year person-level weight (PERWT19F) 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 2019. A key person either was 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.

3.2 Details on Person Weight Construction

The person-level weight PERWT19F was developed in several stages. First, person-level weights for Panel 23 and Panel 24 were created separately. The weighting process for each panel included an adjustment for nonresponse over time and calibration to independent population totals. 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 2019 composite weight was then formed by multiplying each weight from Panel 23 by the factor .50 and each weight from Panel 24 by the factor .50. The choice of factors reflected the relative sample sizes of the two panels, helping to limit the variance of estimates obtained from pooling the two samples. The composite weight was raked to the same set of CPS-based control totals. When the poverty status information derived from income variables became available, a final raking was undertaken, establishing control figures reflecting poverty status rather than educational attainment. Thus, control totals were 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.

3.2.1 MEPS Panel 23 Weight Development Process

The person-level weight for MEPS Panel 23 was developed using the 2018 full year weight for an individual as a "base" weight for survey participants present in 2018. For key, in-scope members who joined an RU some time in 2019 after being out-of-scope in 2018, the initially assigned person-level weight was the corresponding 2018 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 2019 for key, responding persons in-scope on December 31, 2019. These control totals were derived by projecting forward the population distribution obtained from the March 2019 CPS to reflect the December 31, 2019 estimated population total (estimated based on Census projections for January 1, 2020). Variables used for person-level raking included: 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. The final weight for key, responding persons who were not in-scope on December 31, 2019 but were in-scope earlier in the year was the person weight after the nonresponse adjustment.

Note that the 2018 full-year weight that was used as the base weight for Panel 23 was derived as follows; adjustment of the MEPS Round 1 weight for nonresponse over the remaining data collection rounds in 2018; and raking the resulting nonresponse adjusted weight to December 2018 population control figures.

It should also be noted that rather than projecting the March 2019 CPS population distribution estimates forward, the standard approach for MEPS has been to scale back from the following year's CPS estimates. In this case, it would have been the March 2020 CPS estimates. However, there was evidence that the onset of the Covid-19 pandemic in March 2020 in the U.S. affected estimates associated with income and education (Rothbaum & Bee, 2020). Since education was planned as one of the variables to be used for raking, it was decided to use the 2019 March CPS data to establish the population estimates for the Full Year (FY) 2019 weights.

3.2.2 MEPS Panel 24 Weight Development Process

The person-level weight for MEPS Panel 24 was developed using the 2019 MEPS Round 1 person-level weight as a "base" weight. 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 included an adjustment for nonresponse over the remaining data collection rounds in 2019 as well as raking to the same population control figures for December 2019 used for the MEPS Panel 23 weights for key, responding persons in-scope on December 31, 2019. The same six variables employed for Panel 23 raking (educational attainment of the reference person, census region, MSA status, race/ethnicity, sex, and age) were used for Panel 24 raking. Again, the final weight for key, responding persons who were not in-scope on December 31, 2019 but were in-scope earlier in the year was the person weight after the nonresponse adjustment.

Note that the MEPS Round 1 weights for Panel 24 incorporated the following components: the original household probability of selection for the NHIS and for the NHIS sample reserved for MEPS; adjustment for NHIS nonresponse; the probability of selection of NHIS responding households for MEPS; an adjustment for nonresponse at the dwelling unit level for Round 1; and poststratification to U.S. civilian noninstitutionalized population estimates at the family and person level obtained from the corresponding March CPS databases.

3.2.3 The Final Weight for 2019

The final raking of those in-scope at the end of the year has been described above. In addition, the composite weights of two groups of persons who were out-of-scope on December 31, 2019 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, 1999-2018 on CDC

WONDER Online Database, released in 2020, 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, 2019 is 323,833,996 (PERWT19F>0 and INSC1231=1). The sum of the person-level weights across all persons assigned a positive person-level weight is 327,396,693.

3.2.4 Coverage

The target population for MEPS in this file is the 2019 U.S. civilian noninstitutionalized population. However, the MEPS sampled households are a subsample of the NHIS households interviewed in 2017 (Panel 23) and 2018 (Panel 24). 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) or after 2018 (Panel 24) 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 only a small segment of the MEPS target population.

3.3 Using MEPS Data for Trend Analysis

MEPS began in 1996, and the utility of the survey for analyzing health care trends expands with each additional year of data; however, there are a variety of methodological and statistical considerations when examining trends over time using MEPS. Tests of statistical significance should be conducted to assess the likelihood that observed trends may be 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.

In 2013 MEPS survey operations introduced an effort to obtain more complete information about health care utilization from MEPS respondents with full implementation in 2014. This effort resulted in improved data quality and a reduction in underreporting in the second half of 2013 and throughout 2014. Respondents tended to report more visits, especially non-physician visits, by sample members and the new approach appeared particularly effective among those subgroups with relatively large numbers of visits, such as the elderly, Medicare beneficiaries, and people with multiple chronic conditions, disabilities, or poor health. Reported spending on visits also tended to increase, especially for such subgroups.

The aforementioned change 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 major 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.

There are also statistical factors to consider in interpreting trend analyses. Looking at changes over longer periods of time can provide a more complete picture of underlying trends. Analysts may wish to consider using techniques to evaluate, smooth, or stabilize analyses of trends using MEPS data such as comparing pooled time periods (e.g. 1996-97 versus 2011-12), 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, 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

4.1 Developing Event-Level Estimates

The data in this file can be used to develop national 2019 event-level estimates for the U.S. civilian noninstitutionalized population on office-based medical provider visits as well as expenditures, and sources of payment for these visits. Estimates of total visits are the sum of the weight variable (PERWT19F) across relevant event records while estimates of other variables must be weighted by PERWT19F to be nationally representative. The tables below contain event-level estimates for selected variables.

Selected Event-Level Estimates

Office-Based Visits

Estimate of Interest	Variable Name	Estimate (SE)	Estimate Excluding Zero Payment Events (SE)*
Total number of office-based medical provider visits (in millions)	PERWT19F	2,130.3 (64.99)	2,094.0 (63.60)
Total number of in-person visits to doctor (SEEDOC_M18=1, in millions)	PERWT19F	1,035.1 (31.22)	1,018.7 (30.83)
Proportion of office-based medical provider visits with expenditures > 0*	OBXP19X	0.983 (0.0012)	----------

Office-Based Expenditures

Estimate of Interest	Variable Name	Estimate (SE)	Estimate Excluding Zero Payment Events (SE)*
Mean total payments per visit (all sources)	OBXP19X	$240 ($4.0)	$245 ($4.1)
Mean out-of-pocket payment per visit	OBSF19X	$42 ($1.5)	$43 ($1.6)
Mean proportion of total expenditures paid by private insurance per visit	OBPV19X/ OBXP19X	----------	0.376 (0.0075)

Office-Based Expenditures: Physician Visits (SEEDOC_M18 = 1)

Estimate of Interest	Variable Name	Estimate (SE)	Estimate Excluding Zero Payment Events (SE)*
Mean total payments per visit where person saw medical doctor	OBXP19X	$287 ($6.5)	$291 ($6.6)

Estimate of Interest	Variable Name	Estimate (SE)	Estimate Excluding Zero Payment Events (SE)*
Mean out-of-pocket payment per visit where person saw medical doctor	OBSF19X	$44 ($2.2)	$45 ($2.3)
Mean proportion of total expenditures per visit paid by private insurance where person saw medical doctor	OBPV19X/ OBXP19X	----------	0.383 (0.0072)

* Zero payment events can occur in MEPS for the following reasons: (1) the visit was covered under a flat fee arrangement (flat fee payments are included only on the first event covered by the arrangement), (2) there was no charge for a follow-up visit, (3) the provider was never paid directly for services provided by an individual, insurance plan, or other source, (4) the charges were included in another bill, or (5) the event was paid through government or privately funded research or clinical trials.

4.2 Person-Based Estimates for Office-Based Visits

To enhance analyses of office-based visits, analysts may link information about office-based visits by sample persons in this file to the annual full year consolidated file (which has data for all MEPS sample persons), or conversely, link person-level information from the full year consolidated file to this event-level file (see Section 5.0 below for more details). Both this file and the full year consolidated file may be used to derive estimates for persons with office-based care and annual estimates of total expenditures. However, if the estimate relates to the entire population, this file cannot be used to calculate the denominator, as only those persons with at least one office-based event are represented on this data file. Therefore, the full year consolidated file must be used for person-level analyses that include both persons with and without office-based care.

4.3 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 minus 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, flat fee, and zero expenditures) are described in Section 2.5.6.

4.4 Variance Estimation (VARSTR, VARPSU)

The MEPS is based on 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 5.4.2).

4.4.1 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. Retaining this numbering system permits analysts interested in pooling MEPS data across several years to do so without incurring logistical issues related to variance estimation while permitting the establishment of useful estimates of variability for estimates based on data obtained over several years of data collection.

To obtain appropriate standard errors when pooling MEPS data across multiple years, 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. However, starting in 2002, the annual MEPS public use files were released with a common variance structure that allows users to pool data from 2002 and forward.

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 or later, use the variance strata numbering as is.

2. When pooling any year from 1996 to 2001 with any year from 2002 or later, use the pooled linkage public use file HC-036 that contains the proper variance structure. The HC-036 file is updated every year to allow pooling of any year prior to 2002 with any year from 2002 up to the latest year. Further details on the HC-036 file can be found in the public use documentation of the HC-036 file.

4.4.2 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 a person-level file (e.g., MEPS 2019 Full Year Consolidated File) expands the scope of potential estimates. For example, to estimate the total number of office-based medical provider visits of persons with specific demographic characteristics (such as age, race, sex, and education), population characteristics from a person-level file need to be merged onto the office-based medical provider visits file. This procedure is illustrated below. The MEPS 2019 Appendix File, HC-213I, provides additional detail on how to merge MEPS data files.

1. Create data set PERSX by sorting the 2019 Full Year Consolidated File by the person identifier, DUPERSID. Keep only variables to be merged onto the office-based medical provider visits file and DUPERSID.

2. Create data set OBMP by sorting the office-based medical provider visits file by person identifier, DUPERSID.

3. Create final data set NEWOBMP by merging these two files by DUPERSID, keeping only records on the office-based medical provider visits 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=OBMP;
    BY DUPERSID;
RUN;

DATA NEWOBMP;
    MERGE OBMP (IN=A) PERSX(IN=B);
    BY DUPERSID;
    IF A;
RUN;
```

5.2 Linking to the Prescribed Medicines File

The prescribed medicines-event link (RXLK) file provides a link from the MEPS event files to the Prescribed Medicines Event File. When using the RXLK, data users/analysts should keep in mind that one office-based visit can link to more than one prescribed medicine record. Conversely, a prescribed medicine event may link to more than one office-based visit or different types of events. 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. For detailed linking examples, including SAS code, data users/analysts should refer to the MEPS 2019 Appendix File, HC-213I.

5.3 Linking to the Medical Conditions File

The condition-event link (CLNK) file provides a link from MEPS event files to the 2019 Medical Conditions File. When using the CLNK, data users/analysts should keep in mind that (1) conditions are household-reported, (2) there may be multiple conditions associated with an office-based medical provider visit, and (3) a condition may link to more than one office-based medical provider visit or any other type of visit. Users should also note that not all office-based medical provider visits link to the condition file.

References

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.*

Fay, R.E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. *Proceedings of the Survey Research Methods Sections, ASA*, 212-217.

Monheit, A.C., Wilson, R., and Arnett, III, R.H. (Editors). *Informing American Health Care Policy*. (1999). 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.

Shah, B.V., Barnwell, B.G., Bieler, G.S., Boyle, K.E., Folsom, R.E., Lavange, L., Wheeless, 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.

Zuvekas, S.H. and J.W. Cohen. A guide to comparing health care expenditures in the 1996 MEPS to the 1987 NMES. *Inquiry*. 2002 Spring;39(1):76-86. doi: 10.5034/inquiryjrnl_39.1.76. PMID: 12067078.

D. Variable-Source Crosswalk

FOR MEPS HC-213G: 2019 OFFICE-BASED MEDICAL PROVIDER VISITS

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
FFEEIDX	Flat fee ID	CAPI derived
MPCELIG	MPC eligibility flag	Constructed
MPCDATA	MPC data flag	Constructed

Medical Provider Visits Variables

Variable	Description	Source
OBDATEYR	Event date – year	CAPI derived
OBDATEMM	Event date – month	CAPI derived
SEEDOC_M18	Did P talk to MD this visit	MV10
DRSPLTY_M18	MVIS doctor's specialty	MV20
MEDPTYPE_M18	Type of med person P talked to on visit dt	MV30
DOCATLOC	Any MD work at location where P saw prov	MV40
VSTCTGRY	Best category for care P recv on visit dt	MV50
VSTRELCN_M18	This visit related to spec cond	MV60
LABTEST_M18	This visit did P have lab tests	MV90
SONOGRAM_M18	This visit did P have sonogram or ultrsd	MV90
XRAYS_M18	This visit did P have x-rays	MV90
MAMMOG_M18	This visit did P have a mammogram	MV90
MRI_M18	This visit did P have an MRI/Catscan	MV90

Variable	Description	Source
EKG_M18	This visit did P have an EKG, EEG or ECG	MV90
RCVVAC_M18	This visit did P receive a vaccination	MV90
SURGPROC	Was surg proc performed on P this visit	MV80
MEDPRESC	Any medicines prescribed for P this visit	MV110

Flat Fee Variables

Variable	Description	Source
FFOBTYPE	Flat fee bundle	Constructed
FFBEF19	Total # of visits in FF before 2019	FF50
FFTOT20	Total # of visits in FF after 2019	FF60

Imputed Expenditure Variables

Variable	Description	Source
OBSF19X	Amount paid, self/family (imputed)	CP Section (Edited)
OBMR19X	Amount paid, Medicare (imputed)	CP Section (Edited)
OBMD19X	Amount paid, Medicaid (imputed)	CP Section (Edited)
OBPV19X	Amount paid, private insurance (imputed)	CP Section (Edited)
OBVA19X	Amount paid, Veterans/CHAMPVA (imputed)	CP Section (Edited)
OBTR19X	Amount paid, TRICARE (imputed)	CP Section (Edited)
OBOF19X	Amount paid, other federal (imputed)	CP Section (Edited)
OBSL19X	Amount paid, state & local government (imputed)	CP Section (Edited)
OBWC19X	Amount paid, workers' compensation (imputed)	CP Section (Edited)
OBOT19X	Amount paid, other insurance (imputed)	CP Section (Edited)
OBXP19X	Sum of OBSF19X – OBOT19X (imputed)	Constructed
OBTC19X	Household reported total charge (imputed)	CP Section (Edited)
IMPFLAG	Imputation status	Constructed

Weight Variables

Variable	Description	Source
PERWT19F	Expenditure file person weight, 2019	Constructed
VARSTR	Variance estimation stratum, 2019	Constructed
VARPSU	Variance estimation PSU, 2019	Constructed