MEPS HC 248G:
2023 Office-Based Medical Provider Visits

June 2025

Table of Contents

A. Data Use Agreement

Individual identifiers have been removed from the micro-data contained in these files. Nevertheless, under Sections 308 (d) and 903 (c) of the Public Health Service Act (42 U.S.C. 242m and 42 U.S.C. 299 a-1), data collected by the Agency for Healthcare Research and Quality (AHRQ) and/or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which they were supplied; any effort to determine the identity of any reported cases is prohibited by law.

Therefore in accordance with the above referenced Federal Statute, it is understood that:

1. No one is to use the data in this dataset in any way except for statistical reporting and analysis; 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 dataset with individually identifiable records from any datasets other than the Medical Expenditure Panel Survey or the National Health Interview Survey. Furthermore, linkage of the Medical Expenditure Panel Survey and the National Health Interview Survey may not occur outside the AHRQ Data Center, NCHS Research Data Center (RDC) or the U.S. Census RDC network.

By using these data you signify your agreement to comply with the above stated statutorily based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates Title 18 part 1 Chapter 47 Section 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

The Agency for Healthcare Research and Quality requests that users cite AHRQ and the Medical Expenditure Panel Survey as the data source in any publications or research based upon these data.

B.   Background

1.0   Household Component

The Medical Expenditure Panel Survey (MEPS) provides nationally representative estimates of health care use, expenditures, sources of payment, and health insurance coverage for the U.S. civilian 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 care. Estimates can be produced for individuals, families, and selected population subgroups. The panel design of the survey includes five rounds of interviews covering 2 full calendar years. Information about each household member is collected through computer-assisted personal interviewing (CAPI) technology, and the survey builds on this information from interview to interview. All data for a sampled household are reported by a single household respondent.

The MEPS HC was initiated in 1996. Each year, a new panel of sample households is selected. Because the data collected are comparable to those from earlier medical expenditure surveys conducted in 1977 and 1987, it is possible to analyze long-term trends. Historically, each annual MEPS HC sample consists of up to 15,000 households. Data can be analyzed at the person, the family, or the event level. Data must be weighted to produce national estimates.

The set of households selected for each panel of the MEPS HC is a subsample of households participating in the previous year's National Health Interview Survey (NHIS) conducted by the National Center for Health Statistics (NCHS). The NHIS sampling frame provides a nationally representative sample of the U.S. civilian noninstitutionalized population. In 2006, the NCHS implemented a new sample design for the NHIS, to include households with Asian persons in addition to households with Black and Hispanic persons in the oversampling of minority populations. In 2016, NCHS introduced another sample design that discontinued the oversampling of these minority groups.

2.0   Medical Provider Component

When the household CAPI interview is completed and permission is obtained from the sample members to contact their medical provider(s), a sample of these providers is contacted by telephone to obtain information that household respondents cannot accurately provide. This part of the MEPS is called the Medical Provider Component (MPC), and it collects information on dates of visits, diagnosis and procedure codes, and charges and payments. The Pharmacy Component (PC), a subcomponent of the MPC, does not collect data on charges or on diagnosis and procedure codes, but it does collect detailed information on drugs, including the National Drug Code (NDC) and medicine name, as well as amounts of payment. The MPC is not designed to yield national estimates. It is primarily used as an imputation source to supplement/replace household reported expenditure information.

3.0 Survey Management and Data Collection

MEPS HC and MPC data are collected under the authority of the Public Health Service Act. The MEPS HC data are collected under contract with Westat, Inc. and the MEPS MPC data are collected under contract with Research Triangle Institute. Datasets and summary statistics are edited and published in accordance with the confidentiality provisions of the Public Health Service Act and the Privacy Act. The NCHS provides consultation and technical assistance.

As soon as the MEPS data are collected and edited, they are released to the public in stages of microdata files and tables via the MEPS website and datatools.ahrq.gov.

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 2023 MEPS HC and MPC. It was released as an ASCII file (with related SAS, Stata, SPSS, and R programming statements and data user information) and as a SAS dataset, SAS transport file, Stata dataset, and Excel file. The 2023 Office-Based Medical Provider Visits public use file (hereafter referred to as the OB PUF) provides detailed information on office-based provider visits for a nationally representative sample of the U.S. civilian noninstitutionalized population. Data from the OB PUF can be used to make estimates of office-based provider utilization and expenditures for calendar year 2023. The file contains 45 variables and has a logical record length of 211 with an additional 2-byte carriage return/line feed at the end of each record. This PUF consists of MEPS survey data obtained in the 2023 portion of Round 3, and all of Rounds 4 and 5 for Panel 27 and Rounds 1, 2 and the 2023 portion of Round 3 for Panel 28 (i.e., the rounds for the MEPS panels covering calendar year 2023), as illustrated below.
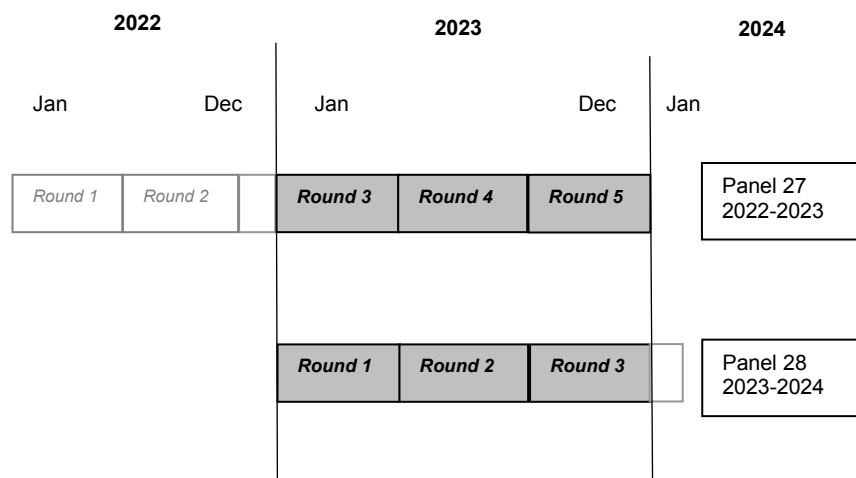
Each record in this PUF represents a unique office-based provider event; that is, an office-based provider event reported by the household respondent. Office-based events reported in Panel 27 Round 5 and Panel 28 Round 3 and known to have occurred after December 31, 2023 are not included in this PUF. Utilization counts of office-based provider visits are based entirely on household reports. Information from the MPC is used to supplement expenditure payment data in the office-based provider file, reported by the household, and does not affect use estimates.

Data from this event PUF can be merged with other 2023 MEPS HC PUFs for purposes of appending person-level data such as demographic characteristics or health insurance coverage to each office-based provider visit record in the current file.

This PUF can also be used to construct summary variables of expenditures, sources of payment, and related aspects of office-based provider visits for calendar year 2023. Aggregate annual person-level information on the use of office-based providers and other health services is provided in the Full Year 2023 Consolidated Public Use File (hereafter referred to as the Consolidated PUF), where each record represents a MEPS sampled person.

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

- Data File Information (Section 2.0)

- Survey Sample Information (Section 3.0)

- Strategies for Estimation (Section 4.0)

- Merging/Linking MEPS Data Files (Section 5.0)

- Variable-Source Crosswalk (Section D)

Any variables not found in this PUF but released in previous years' PUFs may have been excluded because they contained only missing data.

For more information on the MEPS HC sample design, see Chowdhury et al. (2019). Copies of the HC and the MPC survey instruments used to collect the information in this OB PUF are available in the *Survey Questionnaires* section of the MEPS website.

2.0   Data File Information

The 2023 OB PUF consists of one event-level file. The PUF contains characteristics associated with the office-based (OB) event and imputed expenditure data.

This OB PUF contains 135,096 office-based provider event records; of these records, 133,735 are associated with persons having a positive person-level weight (PERWT23F). This PUF 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 2023. Office-based provider visits known to have occurred before January 1, 2023 or after December 31, 2023 are not included in this PUF. Some household members may have multiple events and thus will be represented in multiple records in this PUF. Other household members may have had no events reported and thus will have no records in this PUF. These data were collected during the 2023 portion of Round 3, and all of Rounds 4 and 5 for Panel 27 as

well as Rounds 1, 2, and the 2023 portion of Round 3 for Panel 28 of the MEPS HC. The persons represented in this PUF had to meet either a) or b):

a) Be classified as Key in-scope persons who responded for their entire period of 2023 eligibility (i.e., persons with a positive 2023 full-year person-level weight, PERWT23F > 0), or

b) Be an eligible member of a family whose Key in-scope members have a positive person-level weight (PERWT23F > 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 (FAMWT23F >0). Note that FAMIDYR and FAMWT23F are variables in the 2023 Consolidated PUF.

Persons with no office-based medical provider visit events for 2023 are not included in this event-level OB PUF but are represented in the person-level 2023 Full Year Population Characteristics PUF.

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 characteristics or health insurance coverage to each event record, data from this PUF can be merged with 2023 MEPS HC person-level data (i.e., Consolidated PUF or Population Characteristics PUF) using the person identifier, DUPERSID. The office-based medical provider visit events can also be linked to the MEPS 2023 Medical Conditions Public Use File (hereafter referred to as the Conditions PUF). Please see Section 5.0 or the Appendix to MEPS 2023 Event Files (hereafter referred to as the Appendix PUF) for details on how to merge MEPS data files.

2.1   Codebook Structure

For most variables in this OB PUF, 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 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 corresponds to a unique person and the office-based medical provider visit event identifier corresponds to a unique event.

2.2 Reserved Codes

This OB PUF contains several reserved code values.

Table 1

Reserved Code Values and Definitions

Value	Label	Definition
-1	Inapplicable	Question was not asked due to skip pattern
-7	Refused	Question was asked and respondent refused to answer question
-8	Don't know	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 Cannot be Computed (-15) is assigned to MEPS constructed variables when there was not enough information from the instrument to calculate the constructed variables. Not having enough information is often the result of skip patterns in the data or of missing information stemming from the responses Refused (-7) or Don't Know (-8). Note that, in addition to Don't Know, reserved code -8 also includes cases for which the information from the question was not ascertained.

Generally, values of -1, -7, -8, and -15 for non-expenditure variables have not been edited in this PUF. Analysts who would like to recode these values can find 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 dataset (although the data are also being provided in an Excel file, a Stata dataset, a SAS dataset, and a SAS transport file) and provides the programming identifiers for each variable.

Table 2

Programming Identifiers for Each Variable in the OB PUF

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, the variable names reflect the content of the variable. All imputed/edited variables end with an "X".

As the collection, universe, or categories of variables were altered, some variable names have been appended with "_Myy", where "yy" indicates the collection year in which the alterations were made. Such alterations are described in detail throughout this document.

2.4.1   Variable-Source Crosswalk

Variables in this OB PUF were derived from the CAPI or the MPC data collection instrument, or were 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";

2. Variables 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

   - TH - Telehealth section

3.   Variables constructed from multiple questions by 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, and end in an "X", indicating that they were edited/imputed. Please note that imputed means that a series of logical edits, as well as an imputation process to account for missing data, were performed on the variable.

The total sum of payments and the 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	XP - sum of payments
TR - TRICARE	

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

The fifth and sixth characters indicate the year (23). The seventh character being "X", indicates that the variable is edited/imputed.

For example, OBSF23X is the edited/imputed amount paid by self or family for an office-based medical provider visit expenditure incurred in 2023.

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 NHIS. The dwelling unit ID (DUID) is a 7-digit number consisting of a 2-digit panel number followed by a 5-digit random number assigned after the case was sampled for MEPS. A 3-digit person number (PID) uniquely identifies each person within the DU. The variable DUPERSID is the combination of 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 2023 Population Characteristics PUF.

Record Identifiers (EVNTIDX, FFEEIDX)

EVNTIDX uniquely identifies each office-based medical provider visit event (i.e., each record in this PUF) and is the variable required for linking office-based medical provider visit events to the data file containing details on conditions (MEPS 2023 Medical Conditions PUF). 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: Merging/Linking MEPS Data Files, or the MEPS 2023 Appendix PUF, HC 248I.

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 HC. 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.

Round Indicator (EVENTRN)

EVENTRN indicates the round in which the office-based event was reported. Please note: Rounds 3 (partial), 4, and 5 are associated with MEPS data collected from Panel 27. Likewise Rounds 1, 2, and 3 (partial) are associated with MEPS data collected from Panel 28.

Panel Indicator (PANEL)

PANEL is a constructed variable used to specify the panel number for the person. PANEL will indicate either Panel 27 or Panel 28 for each person in the PUF. Panel 27 is the panel that started in 2022 and Panel 28 is the panel that started in 2023.

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 MPC data were collected for the office-based provider.

2.5.3   Office-Based Medical Provider Visit Variables

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

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.

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 Inapplicable (-1). 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 any medical doctors worked at the visit location (DOCATLOC), the type of care the person received (VSTCTGRY), and whether 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.

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

Services received during the visit included whether 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). All the service and procedure variables are set to -1 for telehealth events. Minimal editing was done across treatments, services, and procedures to ensure consistency across Inapplicable (-1), Refused (-7),

Don't Know (-8), and No Services Received (95) values. Due to design changes, beginning in 2017, EEG was combined with EKG.

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.

Telehealth (TELEHEALTHFLAG - VISITTYPE)

The Telehealth (TH) module is asked of all events tagged as TH events by the respondent. As part of the TH module, a question asks whether the provider or facility is owned or operated by a hospital. Post-collection, the response to this question is used to reclassify all TH events as either OB or OP. The TH module items are designed to align with the existing OB and OP items to easily allow for reclassifying the event type. All events initially reported as TH also have a categorical variable, VISITTYPE, which indicates whether the visit was over the phone, through real-time video, or some other way.

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 in this PUF. For information on the ICD-10-CM condition codes and associated CCSR codes, see the Conditions PUF.

2.5.5   Flat Fee Variables (FFEEIDX, FFOBTYPE)

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. An example is an obstetrician's fee covering a normal delivery, and the associated pre- and post-natal care. A flat fee group is a set of medical services (i.e., events) that are covered under the same flat fee payment. The flat fee groups represented in the OB PUF include flat fee groups where at least one of the health care events, as reported by the HC respondent, occurred during 2023. 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.

It is important to note that certain flat fee bundle types reported by household respondents were identified as having a high likelihood of being simple events misidentified as bundle events. To

address this, starting in 2021, HC-reported flat fee bundles were considered as flat fees if the bundle consisted only of dental events.

Other HC-reported bundles were not allowed as flat fee bundles, and events in these bundles were treated as simple events. HC-reported bundles that included a mix of emergency room and hospitalization events were treated as linked events. All emergency room expenditures were combined with hospital inpatient expenditures. However, provider-reported flat fees were processed in a similar way to prior years.

Flat Fee Variable Descriptions

Flat Fee ID (FFEEIDX)

As noted in "Record Identifiers," the variable FFEEIDX uniquely identifies all events that are part of the same flat fee group for a person. On any 2023 event PUF, every event that is part of a specific flat fee group has the same value for FFEEIDX. Note that prescribed medicine and home health events are never included in a flat fee group and the variable FFEEIDX is not included in those event PUFs.

Flat Fee Type (FFOBTYPE)

FFOBTYPE indicates whether the 2023 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 Inapplicable (-1).

Counts of Flat Fee Events that Cross Years

Starting in 2021, HC-reported flat fee bundles were treated as flat fees only if the bundle consisted exclusively of dental events. Other HC-reported bundles were not recognized as flat fee bundles; instead, the events within those bundles were treated as simple events. Because flat fee bundles can span multiple years, 2023 is the first year in which FFBEFYY and FFTOTYY contain no valid values, so these variables have been removed from the PUF.

Caveats of Flat Fee Groups

Analysts should note that flat fee payments are common in the OB PUF. In the 2023 PUF there are 543 events that are identified as being part of a flat fee payment group. To correctly identify all events that are part of a flat fee group, the analyst should link all MEPS events 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 2023, but the remaining visits that were part of this flat fee group occurred in 2024. In this case, the 2023 flat fee group represented in this PUF would consist of one event (the stem). The 2024 leaf events that are part of this flat fee group are not represented in this PUF. Similarly, the household respondent may have reported a flat fee group where the initial visit began in 2022, but subsequent visits occurred during 2023.

In this case, the initial visit would not be represented in this PUF. This 2023 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. Note that the Variable-Source Crosswalk lists all possible flat fee variables.

2.5.6   Expenditure Data

Definition of Expenditures

Expenditures in this PUF refer to payments 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 from its predecessors: the 1987 NMES and 1977 NMCES surveys where "charges" rather than the 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 associated payments. While charge data are provided in this PUF, 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, or the expenditures defined in the 1987 NMES. For details on expenditure definitions, please refer to Monheit, et al. (1999). AHRQ has developed factors to apply to the 1987 NMES expenditure data to facilitate longitudinal analysis. These factors are published in Zuvekas and Cohen (2002), and also 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 3.5 for more information.

Data Editing and Imputation Methodologies of Expenditure Variables

The expenditure data included in this PUF were derived from both the MEPS HC and 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 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, 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.

A weighted sequential hot-deck procedure was used to impute the missing total charges. This procedure uses survey data from donors to replace missing data while taking into account the donors' weighted distribution in the imputation process to ensure that the weighted distribution of recipients' expenditures reflects the weighted distribution of the donors' expenditures.

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 separate 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.

Capitation Imputation

A 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.

Imputation Flag (IMPFLAG)

IMPFLAG is a six-category variable that indicates if the event contains complete HC or 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

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 2023, all the events that occurred in 2023 will have zero payments. Conversely, if the first event in the flat fee group occurred at the end of 2023, the total expenditure for the entire flat fee group will be on that event, regardless of the number of events it covered after 2023. See Section 2.5.5 for details on the flat fee variables.

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.

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.

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 allowed to remain - the amounts are not moved to Other Private and Other Public categories anymore. The two source of payment categories, Other Private and Other Public, are no longer available.

Office-Based Expenditure Variables (OBSF23X - OBTC23X)

OBSF23X - OBOT23X are the 10 sources of payment. The 10 sources of payment are: self/family (OBSF23X), Medicare (OBMR23X), Medicaid (OBMD23X), private insurance (OBPV23X), Veterans /CHAMPVA (OBVA23X), TRICARE (OBTR23X), other federal sources (OBOF23X), state and local (non-federal) government sources (OBSL23X), Workers'

Compensation (OBWC23X), and other insurance (OBOT23X). OBXP23X is the sum of the 10 sources of payment for the office-based expenditures, and OBTC23X is the total charge.

2.5.7   Rounding

Expenditure variables have been rounded to the nearest penny. Person-level expenditure information released in the Consolidated PUF will be rounded to the nearest dollar. Note that using the event files to create person-level totals will yield slightly different totals than those found in the Consolidated PUF. These differences are due to rounding only. Moreover, in some instances, the number of persons with expenditures in the event PUFs for a particular source of payment may differ from the number of persons with expenditures in the person-level Consolidated PUF 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 MEPS Data

Modification to the MEPS sample design because of the pandemic ended in 2022. Concerns of potential bias due to these modifications should no longer apply to data collected in this PUF. However, like most other surveys, MEPS has been substantially affected by the pandemic. As a result of these changes, potential bias continues to be a concern. One effect of the pandemic is the significantly lower response rates (Section 3.2), and these lower rates might differentially exclude households more likely to experience IP stays. The demographic shifts on MEPS between 2019 and 2022 suggest a more educated, higher-income, older MEPS sample. (For more detail, see Section 3.1 of the 2020 Consolidated PUF, Section 3.1 of the 2021 Consolidated PUF, and Section 3.1.2 of the 2022 Consolidated PUF.)

Analyses undertaken to examine the quality of the MEPS FY 2023 data compare health care utilization and health insurance coverage for the MEPS target population between the panels fielded. These comparisons were undertaken for the full sample and the three age groups of 0-17, 18-64, and 65+. These comparisons found no abnormal differences between the two panels. Analyses across years also suggest a rebound to pre-pandemic utilization levels for most key event types.

The various actions taken in the development of the person-level weights for the MEPS FY 2023 data were designed to limit the potential for response bias. However, evaluations of MEPS data quality in 2020 through 2022 suggest that analysts of the MEPS FY 2023 Population Characteristics PUF should continue to exercise caution when interpreting estimates and assessing analyses based on data collected from these three calendar years. This includes the comparison of such estimates to those of other years and corresponding trend analyses.

3.2   Sample Weight (PERWT23F)

There is a single full-year person-level weight (PERWT23F) assigned to each record for each Key, in-scope person who responded to MEPS for the full period of time that the person was in scope during 2023. A Key person was either a member of a responding NHIS household at the time of the 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 the person is a member of the civilian noninstitutionalized portion of the U.S. population.

3.3   Details on Person Weight Construction

The person-level weight PERWT23F was developed in several stages. First, a person-level weight for Panel 27 was created, including an adjustment for nonresponse over time and raking. The raking involved adjusting to several sets of marginal control totals reflecting 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 (three categories: no degree; high school/GED only or some college; bachelor's or a 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 (0-18, 19-25, 26-34, 35-44, 45-64, and 65 or older). (Note, however, that for confidentiality reasons, the MSA status variables are no longer released for public use.) The person-level weight for Panel 28 was created similarly. Secondly, a composite weight was formed by multiplying each weight from Panel 27 by the factor .40 and each weight from Panel 28 by the factor .60. The choice of factors reflected the relative effective sample sizes of the two panels, helping to limit the variance of estimates obtained from pooling both samples. Weights for the 2023 Population Characteristics PUF were then developed by raking the composite weight to the same set of CPS-based control totals.

The approach for establishing the 2023 Consolidated PUF weight is as follows. When poverty status information derived from MEPS 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, replacing educational attainment with poverty status while retaining the other five raking variables previously indicated. Specifically, control totals based on CPS estimates of 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 age, race/ethnicity, sex, region, and MSA status are used to calibrate weights.

3.3.1 MEPS Panel 27 Weight Development Process

The person-level weight for MEPS Panel 27 was developed by using the 2022 full-year weight as a "base" weight for survey participants present in 2023.

For Key, in-scope members who joined an RU at some time in 2023 after being out of scope in 2022, the initially assigned person-level weight was the corresponding 2022 family weight. The weighting process also included an adjustment for person-level nonresponse over Rounds 4 and 5 as well as raking to the population control figures for December 2023 for Key, responding persons in scope on December 31, 2023. These control totals were derived by scaling back the population distribution obtained from the March 2024 CPS to reflect the December 31, 2023 estimated population total (estimated based on Census projections for January 1, 2024). Variables used for person-level raking included: education of the reference person (no degree, high school/GED only or some college, bachelor's or a 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. (Note, however, that for confidentiality reasons, the MSA status variables are no longer released for public use.) The final weight for Key, responding persons who were not in scope on December 31, 2023 but were in scope earlier in the year was the nonresponse-adjusted person weight without raking.

Note that the 2022 full-year weight that was used as the base weight for Panel 27 was derived using the 2022 MEPS Round 1 weight and reflected adjustment for nonresponse over the remaining data collection rounds in 2022 as well as raking to the December 2022 population control figures.

3.3.2   MEPS Panel 28 Weight Development Process

The person-level weight for Panel 28 was developed using the 2023 Round 1 person-level weight as a "base" weight. The Round 1 weights incorporated the following components: the original household probability of selection for the NHIS and for the NHIS subsample reserved for the MEPS, an adjustment for NHIS nonresponse, the probability of selection for MEPS from the NHIS, an adjustment for nonresponse at the dwelling unit level for Round 1, and raking 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 DU weight served as a "base" weight.

The weighting process also included an adjustment for nonresponse over the remaining data collection rounds in 2023 as well as raking to the same population control figures for December 2023 that were used for the Panel 27 weight for Key, responding persons in scope on December 31, 2023. The same six variables used for Panel 27 raking (education level of the reference person, Census region, MSA status, race/ethnicity, sex, and age) were also used for Panel 28 raking. Similar to Panel 27, the Panel 28 final weight for Key, responding persons who were not in scope on December 31, 2023 but were in scope earlier in the year was the nonresponse-adjusted person weight without raking.

3.3.3   The Final Weight for 2023

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, 2023 were adjusted for expected undercoverage. Specifically, the weights of those who were out of scope on December 31, 2023, but in scope at some time during the year and were 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), Provisional Mortality Statistics, 2018 through Last Week on CDC WONDER Online Database, released in 2024, 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, 2023 is 330,710,135 (PERWT23F >0 and INSC1231=1). The sum of person-level weights across all persons assigned a positive person-level weight is 334,530,273.

3.4 Coverage

The target population associated with MEPS is the 2023 U.S. civilian noninstitutionalized population. However, the MEPS sampled households are a subsample of the NHIS households interviewed in 2021 (Panel 27) and 2022 (Panel 28). New households created after the NHIS interviews for the respective panels and consisting exclusively of persons who entered the target population after 2021 (Panel 27) or after 2022 (Panel 28) are not covered by the 2023 MEPS. Nor are previously out of scope persons who joined an existing household but are not related 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. Those not covered represent a small proportion of the MEPS target population.

3.5 Using MEPS Data for Trend Analysis

For analysts using the MEPS data for trend analysis, we note that there are uncertainties associated with 2020, 2021, and 2022 data quality for reasons discussed throughout Section 3. Evaluations of important MEPS estimates suggest that they are of reasonable quality. Nevertheless, analysts are advised to exercise caution in interpreting these estimates, particularly in terms of trend analyses, since access to health care was substantially affected by the pandemic, as were related factors such as health insurance and employment status for many persons.

The MEPS began in 1996, and the utility of the survey for analyzing health care trends expands with each additional year of data; however, when examining trends over time using the MEPS, the length of time being analyzed should 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 the MEPS methodology.

With respect to methodological considerations, changes in data collection methods, such as interviewer training, were introduced in 2013 to obtain more complete information about health care utilization from MEPS respondents; the changes were fully implemented in 2014. This effort likely resulted in improved data quality and a reduction in underreporting starting in the second half of 2013 and continuing throughout the 2014 full year files; the changes have also had some impact on analyses involving trends in utilization across years. The changes in the NHIS sample design in 2016 and 2018 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 United States. As a result, it can be expected to better cover the full 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 MEPS files were established from data collected in Rounds 1-3 of Panel 22 and Rounds 3-5 of Panel 21, they reflected two instrument designs. 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 easier 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. *Analysts should be aware of the possible impacts of these changes on the data and especially trend analyses that include the year 2018 because of the design transition.*

Process changes, such as data editing and imputation, may also affect trend analyses. For example, analysts should refer to Section 2.5.11: Utilization, Expenditures, and Source of Payment Variables in the Consolidated PUF (HC 251) and, for more detail, to the documentation for the prescription drug file (HC 248A) when analyzing prescription drug spending over time.

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

To smooth or stabilize trend analyses based on the MEPS data, analysts may also wish to consider using statistical techniques 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 the data.

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, conducting numerous statistical significance tests of trends will increase 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 PUF can be used to develop national 2023 event-level estimates for the U.S. civilian noninstitutionalized population for 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 (PERWT23F) across relevant event records while estimates of other variables must be weighted by PERWT23F to be nationally representative. The tables below contain event-level estimates for selected variables.

Table 3

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)	PERWT23F	2,214.2 (68.04)	2,186.1 (67.49)
Total number of in-person visits to doctor (SEEDOC_M18=1, in millions)	PERWT23F	973.6 (28.14)	955.4(27.68)
Proportion of office-based medical provider visits with expenditures > 0*	OBXP23X	0.987 (0.0008)	-------------

* 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.

Table 4

Selected Event-Level Estimates - Office-Based Expenditures

Estimate of interest	Variable name	Estimate (SE)	Estimate excluding zero payment events (SE)
Mean total payments per visit (all sources)	OBXP23X	$285 ($6.1)	$289 ($6.2)
Mean out-of-pocket payment per visit	OBSF23X	$53 ($2.6)	$53 ($2.6)
Mean proportion of total expenditures paid by private insurance per visit	OBPV23X/ OBXP23X	-------------	0.378 (0.0081)

Table 5

Selected Event-Level Estimates - 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	OBXP23X	$347 ($10.9)	$354 ($11.1)
Mean out-of-pocket payment per visit where person saw medical doctor	OBSF23X	$58 ($4.5)	$59 ($4.5)
Mean proportion of total expenditures per visit paid by private insurance where person saw medical doctor	OBPV23X/ OBXP23X	-------------	0.374 (0.0085)

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 Consolidated PUF (which has data for all MEPS sample persons), or conversely, link person-level information from the Consolidated PUF to this event-level PUF (see Section 5.0 below for more details). Both this PUF and the Consolidated PUF may be used to derive estimates for persons with office-based care and annual estimates of total expenditures. However, for estimates that pertain to those who did not receive office-based care as well as those who did (for example, the percentage of adults with at least one office-

based visit during the past year or the mean number of office-based visits in the past year among those 65 or older), this PUF cannot be used. Only those persons with at least one office-based event are represented in this PUF. The Consolidated PUF 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 calculated, 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, flat fee, and zero expenditures) are described in Section 2.5.6.

4.4 Variance Estimation (VARSTR, VARPSU)

To obtain estimates of variability in the MEPS estimates (such as the standard error of sample estimates or corresponding confidence intervals), analysts should consider the complex sample design of the 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 (BRR), 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 PUF does not contain replicate weights, analysts can use the BRR methodology to construct replicate weights to develop variances for more complex estimators (see Section 4.4.2).

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 and all other MEPS PUFs. 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 user documentation for the software.

With the Taylor series linearization method, variance estimation strata and the variance estimation PSUs within these strata must be specified. The variables VARSTR and VARPSU in this OB PUF 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 software packages will provide estimated standard errors appropriate for assessing the variability

of the MEPS estimates. Note 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 see at least 100 degrees of freedom associated with the estimated standard errors for national estimates based on this MEPS database.

Before 2002, the 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. Beginning with the 2002 Point-in-Time PUF, the approach changed with the intention that variance strata and PSUs would be developed to be compatible with all future PUFs until the NHIS design changed. Thus, when pooling data across years 2002 through Panel 11 of the 2007 files, analysts can use the variance strata and PSU variables provided without modifying them for variance estimation purposes for estimates covering multiple years of data. There are 203 variance estimation strata, each stratum with either two or three variance estimation PSUs.

Beginning in Panel 12 of the 2007 files, a new set of variance strata and PSUs was 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. Therefore, there are a total of 368 (203+165) variance strata in the 2007 Population Characteristics PUF, as it consisted 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 same NHIS design, there are only 165 variance strata. These 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 if data are pooled for several years.

The NHIS sample design was changed again in 2016, effectively changing the MEPS design beginning with calendar year 2017. Beginning with Panel 22 of the 2017 files, a new set of variance strata and PSUs were developed. There are 117 variance strata with either two or three variance estimation PSUs per stratum. Therefore, there are a total of 282 (165+117) variance strata in the 2017 Population Characteristics PUF, as it consisted of two panels that were selected under two independent NHIS sample designs. To make the pooling of data across multiple years of the MEPS more straightforward, the numbering system for the variance strata was changed. The strata associated with the new design are numbered from 2001 to 2117.

The NHIS sample design was further modified in 2018, so the MEPS variance structure for the 2019 Population Characteristics PUF was also modified, reducing the number of variance strata to 105. Consistency was maintained with the prior structure in that the 2019 variance strata were also numbered within the range of values from 2001 to 2117, although there are now gaps in the values assigned within this range. Because of the modification, each stratum could contain up to 5 variance estimation PSUs.

For Panel 26 in the 2021 and 2022 Population Characteristics PUFs, an additional NHIS sample was used for the MEPS to account for increasing nonresponse during the pandemic (as discussed in Section 3.1). The additional sample was assigned to the existing variance strata, so the 2021 and 2022 Population Characteristics PUFs continued to have 105 variance strata, numbered 2001-2117, with a few gaps in the values in that range. In many cases, the additional sample was

assigned to new variance estimation PSUs, so in the 2021 and 2022 Population Characteristics PUFs, each stratum contained up to eight variance estimation PSUs.

Additional NHIS samples were no longer needed in 2023, leading to fewer variance estimation PSUs than in the 2021 and 2022 Population Characteristics PUFs. The 2023 Population Characteristics PUF continues to have 105 variance strata, numbered 2001-2117, with a few gaps in the values in that range. Each stratum contains up to six variance estimation PSUs.

Some analysts may be interested in pooling data across multiple years of MEPS data. When doing so, analysts should note that, to obtain appropriate standard errors, it is necessary to specify a common variance structure. Before 2002, each annual PUF was released with a variance structure unique to the particular MEPS sample in that year. Starting in 2002, the annual PUFs were released with a common variance structure that allowed analysts to pool data from 2002 through 2018. However, analysts can no longer do this routinely because the variance structure had to be modified beginning with 2019.

To ensure that variance strata are identified appropriately for variance estimation purposes when pooling MEPS data across several years, analysts 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) the year 2019 and beyond with any earlier year, use the pooled linkage PUF HC-036, which 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 are included in the public use documentation of the HC-036 file.

4.4.2   Balanced Repeated Replication Method

BRR replicate weights are not provided in this MEPS PUF for the purposes of variance estimation. However, a file containing a BRR replication structure is made available so that analysts 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 for computing variances of complex nonlinear estimators for which a Taylor linear form is neither easy to derive nor available in commonly used software. For instance, it is not possible to calculate the variances of a median or the ratio of two medians by using the Taylor linearization method. For these types of estimators, analysts can 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 by 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 1996-2021 Replicates for Variance Estimation File contains the information necessary to construct the BRR replicates. It includes 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 an analysis of MEPS data pooled across years, the BRR replicates can be formed in the same way by using the HC-036, MEPS 1996-2021 Pooled Linkage Variance Estimation File. For more information about creating BRR replicates, analysts 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 PUF can be used alone or in conjunction with other PUFs for different analytic purposes. Merging characteristics of interest from other MEPS PUFs expands the scope of potential estimates. For example, the medical event PUFs can be merged with the person-level Consolidated PUF to calculate event-level estimates for persons with specific characteristics (e.g., age, race, sex, and education).

Most of the event PUFs can also be linked to the Conditions PUF by using the condition-event link (CLNK) PUF. When using the CLNK PUF, analysts should keep in mind that (1) conditions are household reported, (2) there may be multiple conditions associated with a medical event, (3) one condition may link to more than one event, and (4) not all medical events link to the Conditions PUF.

In addition to linking to other MEPS PUFs, each MEPS panel can also be linked back to the previous year's NHIS public use files. This is because the set of households selected for MEPS is a subsample of those participating in the NHIS. For information on obtaining MEPS/NHIS link files please see the MEPS website.

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. (2019, January). Methodology Report #33. Rockville, MD: Agency for Healthcare Research and Quality.

Cohen, S.B. (1996). The redesign of the Medical Expenditure Panel Survey: A component of the DHHS survey integration plan. *Proceedings of the Council of Professional Associations on Federal Statistics Seminar on Statistical Methodology in the Public Service.*

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. Hyattsville, MD: National Center for Health Statistics.

Fay, R.E. (1989). Theory and application of replicate weighting for variance calculations. *Proceedings of the Survey Research Methods Sections of the American Statistical Association*, 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), 2118-2121.

Monheit, A.C., Wilson, R., and Arnett, III, R.H. (Eds.). (1999). *Informing American health care policy*. Jossey-Bass Inc.

Rothbaum, J. & Bee, A. (2021, May 3). *Coronavirus infects surveys, too: Survey nonresponse bias and the coronavirus pandemic*. Washington, DC: U.S. Census Bureau.

Rothbaum, J. & Bee, A. (2022, September 13). *How has the pandemic continued to affect survey response? Using administrative data to evaluate nonresponse in the 2022 Current Population Survey Annual Social and Economic Supplement.* Washington, DC: 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 Institute.

U.S. Census Bureau. *Current Population Survey: 2021 Annual Social and Economic (ASEC) Supplement*. (2021). Washington, DC: Author.

Zuvekas, S.H. & Cohen, J.W. (2002). A guide to comparing health care expenditures in the 1996 MEPS to the 1987 NMES. *Inquiry*. 39(1), 76-86.

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), 2157-2166.

D. Variable-Source Crosswalk

FOR MEPS HC 248G: 2023 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
TELEHEALTHFLAG	Is this a telehealth event	Constructed
VISITTYPE	Type of telehealth visit	TH10

Flat Fee Variables

Variable	Description	Source
FFOBTYPE	Flat fee bundle	Constructed

Imputed Expenditure Variables

Variable	Description	Source
OBSF23X	Amount paid, self/family (imputed)	CP Section (Edited)
OBMR23X	Amount paid, Medicare (imputed)	CP Section (Edited)
OBMD23X	Amount paid, Medicaid (imputed)	CP Section (Edited)
OBPV23X	Amount paid, private insurance (imputed)	CP Section (Edited)
OBVA23X	Amount paid, Veterans/CHAMPVA (imputed)	CP Section (Edited)
OBTR23X	Amount paid, TRICARE (imputed)	CP Section (Edited)
OBOF23X	Amount paid, other federal (imputed)	CP Section (Edited)
OBSL23X	Amount paid, state & local government (imputed)	CP Section (Edited)
OBWC23X	Amount paid, workers' compensation (imputed)	CP Section (Edited)
OBOT23X	Amount paid, other insurance (imputed)	CP Section (Edited)
OBXP23X	Sum of OBSF23X - OBOT23X (imputed)	Constructed
OBTC23X	Household reported total charge (imputed)	CP Section (Edited)
IMPFLAG	Imputation status	Constructed

Weight Variables

Variable	Description	Source
PERWT23F	Expenditure file person weight, 2023	Constructed
VARSTR	Variance estimation stratum, 2023	Constructed
VARPSU	Variance estimation PSU, 2023	Constructed