MEPS HC 239I:
Appendix to MEPS 2022 Event Files
HC 239A – HC 239H

August 2024

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

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 two full calendar years. Additional rounds were added to Panel 24 in 2021 and 2022, covering the third and fourth years respectively, to compensate for the smaller number of completed interviews in later panels. These extra rounds provide data for examining person-level changes in selected variables such as expenditures, health insurance coverage, and health status. Information about each household member is collected through computer-assisted personal interviewing (CAPI) technology, and the survey builds on this information from interview to interview. All data for a sampled household are reported by a single household respondent.

The MEPS HC was initiated in 1996. Each year a new panel of sample households is selected. Because the data collected are comparable to those from earlier medical expenditure surveys conducted in 1977 and 1987, it is possible to analyze long-term trends. Historically, each annual MEPS HC sample consists of approximately up to 15,000 households. Data can be analyzed at the person, the family, or 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 household survey respondents 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, 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, 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 the MEPS Public Use Release HC 239I, which is the Appendix to MEPS releases HC 239A through HC 239H. This release contains the condition-event link file (hereafter referred to as the CLNK PUF), provided in ASCII (with related SAS, SPSS, R, and Stata programming statements and data user information) and as a SAS dataset, a SAS transport file, a Stata dataset, and an Excel file.

This documentation offers a brief overview of the content and structure of the PUF and the accompanying codebook. It contains the following sections:

- Data File Information (Section 2.0)
- Merging/Linking MEPS Data Files (Section 3.0)

For more information on MEPS HC sample design see Chowdhury et al. (2019). For information on the MEPS MPC design, see RTI International (2024). A copy of the survey instrument used to collect the information in this PUF, is available on the MEPS website.

2.0 Data File Information

This CLNK PUF contains variables for linkage of the MEPS 2022 event-level data files. The H239IF1 or CLNK PUF, is used for linking the MEPS Conditions PUF with the MEPS event PUFs. The CLNK PUF contains 6 variables and has a logical record length of 71 with an additional 2-byte carriage return/line feed at the end of each record.

2.1 Codebook Format

The codebook describes an ASCII dataset (although the data are also being provided in a SAS dataset, a SAS transport file, a Stata dataset, and an Excel file), and provides programming identifiers for each variable.

Table 1

Programming Identifiers for Each Variable in the CLNK PUF

Identifier	Description
Name	Variable name
Description	Variable descriptor
Format	Number of bytes

Identifier	Description
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

2.2 Variable Naming and Source

In general, the variable names reflect the content of the variable. All variables contained in this PUF were derived from the CAPI.

2.3 Contents of Condition-Event Link File (CLNK)

The CLNK PUF contains the variables needed to link each record on the MEPS 2022 Medical Conditions PUF (hereafter referred to as the Conditions PUF), HC 241, with one or more records on the MEPS 2022 event files, HC 239A and HC 239D through HC 239H. Section 3.0 contains additional information on completing this linkage.

The ten-character variable DUPERSID uniquely identifies each person represented on this PUF. The variable DUPERSID is the combination of the variables DUID and PID. All ID variables begin with the 2-digit panel number. There may be more than one record on the CLNK PUF for a specific DUPERSID value.

CONDIDX is the 13-digit ID that uniquely identifies each condition for a person and corresponds to a unique record on the MEPS 2022 Conditions PUF, HC 241. The variable CONDIDX is the combination of the variables DUPERSID and CONDN (see HC 241 for a description of CONDN). The 2-digit panel number is added in the beginning of CONDIDX. There may be more than one record on the CLNK PUF for a specific CONDIDX value.

EVNTIDX is the 16-digit number that uniquely identifies each event for a person and corresponds to a unique record on one of the MEPS 2022 event files, HC 239B through HC 239H. (EVNTIDX is not included on the 2022 Prescribed Medicines event PUF, HC 239A; rather, in this PUF the variable for linking with EVNTIDX on the CLNK PUF is LINKIDX.) There may be more than one record on the CLNK PUF for a specific EVNTIDX value. The 2-digit panel number is added in the beginning of EVNTIDX, and a 2-digit event type number is added to the end. The event type number indicates the type of event record and has been rolled up into the following values:

01 = MVIS - office-based medical provider visit event on MEPS release HC 239G or

OPAT - outpatient department visit event on MEPS release HC 239F or

EROM - emergency room visit event on MEPS release HC 239E or

STAZ - inpatient hospital stay event on MEPS release HC 239D or

HVIS - home health visit event on MEPS release HC 239H

03 = PMED - prescribed medicine event on MEPS release HC 239A

CLNKIDX is the 29-digit number that uniquely identifies each record on the CLNK PUF and is the combination of CONDIDX + EVNTIDX. There is just one record in this PUF for each value of CLNKIDX, i.e., each unique combination of CONDIDX + EVNTIDX.

The variable EVENTYPE indicates the type of event record, and has the following values:

1 = MVIS - office-based medical provider visit event contained on MEPS release HC 239G

2 = OPAT - outpatient department visit event contained on MEPS release HC 239F

3 = EROM - emergency room visit event contained on MEPS release HC 239E

4 = STAZ - inpatient hospital stay event contained on MEPS release HC 239D

7 = HVIS - home health visit event contained on MEPS release HC 239H

8 = PMED - prescribed medicines event contained on MEPS release HC 239A

PANEL is a constructed variable used to specify the panel number for the interview in which the condition was reported. PANEL will indicate either Panel 24, Panel 26, or Panel 27. Panel 24 is the panel that started in 2019, Panel 26 is the panel that started in 2021, and Panel 27 is the panel that started in 2022. The panel number is included as the first two digits of the DUID and DUPERSID.

2.4 ICD-10-CM, CCSR1X, CCSR2X, CCSR3X, and CCSR4X

ICD-10-CM diagnosis codes and Clinical Classification Software Refined (CCSR) codes are both used to group medical conditions into clinically meaningful categories. For the purposes of MEPS, one ICD-10-CM diagnosis code may map to up to four CCSR categories (CCSR1X, CCSR2X, CCSR3X, and CCSR4X) using the v2023.1 release of the CCSR for ICD-10-CM diagnoses. For more information on CCSR, visit the user guide for CCSR.

3.0 Merging/Linking MEPS Data Files

This PUF is intended to be used in conjunction with other PUFs. Specifically, the Conditions PUF (HC 241), the Prescribed Medicines event PUF (HC 239A), and event files HC 239D through HC 239H.

3.1 Limitations/Caveats of the CLNK PUF

When using the CLNK PUF, analysts should keep in mind that (1) conditions are self-reported, (2) there may be multiple conditions associated with an event, and (3) there may be multiple events associated with a condition. Analysts should also note that not all events link to the Conditions PUF.

3.2 National Health Interview Survey

Data from this PUF can be used alone or in conjunction with other PUFs for different analytic purposes. Each MEPS panel can also be linked back to the previous years' NHIS Survey public use data files. For information on obtaining MEPS/NHIS link files please see the MEPS website.

3.3 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 possibly 2022 data quality for reasons discussed in Section 3.0: Survey Sample Information of the Full Year Consolidated PUF document (HC 243). Preliminary evaluations of a set of MEPS estimates of particular importance 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 COVID-19 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 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 on 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 the Section 2.5.11:Utilization, Expenditures, and Sources of Payment Variables in the Full Year Consolidated PUF (HC 243) and, for more detail, to the documentation for the prescription drug file (HC 239A), 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, the use of numerous statistical significance tests of trends will increase the likelihood of concluding that a change has taken place when one has not.

3.4 Longitudinal Analysis

Panel-specific longitudinal files can be downloaded from the data section of the MEPS website. For all three panels (Panel 24, Panel 26, and Panel 27), the longitudinal file comprises MEPS data obtained in all rounds of the panel and can be used to analyze changes over the entire length of the panel. Variables in the file pertaining to survey administration, demographics, employment, health status, disability days, quality of care, patient satisfaction, health insurance, and medical care use and expenditures were obtained from the MEPS Consolidated PUFs from the years covered by that panel.

For more details or to download the data files, please see Longitudinal Weight files at the AHRQ website.

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. Rockville, MD: Agency for Healthcare Research and Quality.

RTI International (2024). *Medical Expenditure Panel Survey Medical Provider Component (MEPS-MPC) Methodology Report 2022 Data Collection*. Rockville, MD. Agency for Healthcare Research and Quality.