

MEPS HC-220I: Appendix to MEPS 2020 Event Files HC-220A - HC-220H

August 2022

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

Agency for Healthcare Research and Quality
Center for Financing, Access, and Cost Trends
5600 Fishers Lane
Rockville, MD 20857
(301) 427-1406

Table of Contents

Section		Page
A.	Data Use Agreement.....	A-1
B.	Background.....	B-1
	1.0 Household Component	B-1
	2.0 Medical Provider Component.....	B-1
	3.0 Survey Management and Data Collection.....	B-2
C.	Technical and Programming Information	C-1
	1.0 General Information.....	C-1
	2.0 Data File Information	C-1
	2.1 Codebook Format	C-1
	2.2 Variable Naming and Source	C-2
	2.3 Contents of File 1: Condition-Event Link File (CLNK).....	C-2
	2.4 Contents of File 2: Prescribed Medicines-Event Link File (RXLK)	C-3
	2.5 ICD-10-CM, CCSR1X, CCSR2X, and CCSR3X.....	C-4
	3.0 Merging/Linking MEPS Data Files	C-5
	3.1 Limitations/Caveats of the CLNK File	C-5
	3.2 Limitations/Caveats of the RXLK File.....	C-5
	3.3 National Health Interview Survey.....	C-5
	3.4 Using MEPS Data for Trend Analysis	C-5
	3.5 Longitudinal Analysis.....	C-7
	References.....	C-8

A. Data Use Agreement

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

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

1. No one is to use the data in this data set in any way except for statistical reporting and analysis; and
2. If the identity of any person or establishment should be discovered inadvertently, then (a) no use will be made of this knowledge, (b) the Director Office of Management AHRQ will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHRQ, and (d) no one else will be informed of the discovered identity; and
3. No one will attempt to link this data set with individually identifiable records from any data sets other than the Medical Expenditure Panel Survey or the National Health Interview Survey. Furthermore, linkage of the Medical Expenditure Panel Survey and the National Health Interview Survey may not occur outside the AHRQ Data Center, NCHS Research Data Center (RDC) or the U.S. Census RDC network.

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

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

B. Background

1.0 Household Component

The Medical Expenditure Panel Survey (MEPS) provides nationally representative estimates of health care use, expenditures, sources of payment, and health insurance coverage for the U.S. civilian noninstitutionalized population. The MEPS Household Component (HC) also provides estimates of respondents' health status, demographic and socio-economic characteristics, employment, access to care, and satisfaction with health care. Estimates can be produced for individuals, families, and selected population subgroups. The panel design of the survey, which includes 5 Rounds of interviews covering 2 full calendar years and two additional rounds in 2020 covering a third year to compensate for the smaller number of completed interviews in Panel 25, provides data for examining person level changes in selected variables such as expenditures, health insurance coverage, and health status. Using computer assisted personal interviewing (CAPI) technology, information about each household member is collected, and the survey builds on this information from interview to interview. All data for a sampled household are reported by a single household respondent.

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

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

2.0 Medical Provider Component

Upon completion of the household CAPI interview and obtaining permission from the household survey respondents, a sample of medical providers are contacted by telephone to obtain information that household respondents cannot accurately provide. This part of the MEPS is called the Medical Provider Component (MPC) and information is collected on dates of visits, diagnosis and procedure codes, charges and payments. The Pharmacy Component (PC), a subcomponent of the MPC, does not collect charges or diagnosis and procedure codes but does collect drug detail information, including National Drug Code (NDC) and medicine name, as well as amounts of payment. The MPC is not designed to yield national estimates. It is primarily used as an imputation source to supplement/replace household reported expenditure information.

3.0 Survey Management and Data Collection

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

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases, micro data files, and tables via the [MEPS website](#).

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857 (301-427-1406).

C. Technical and Programming Information

1.0 General Information

This documentation describes the MEPS Public Use Release HC-220I, which is the Appendix to MEPS releases HC-220A through HC-220H. This release contains two data files, the condition-event link file (CLNK) and the prescribed medicines-event link file (RXLK), both of which are provided in ASCII (with related SAS, SPSS, R, and Stata programming statements and data user information) and SAS data set, SAS transport file, Stata data set, and Excel file versions.

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

- Data File Information
- Merging/Linking MEPS Data Files

For more information on MEPS HC sample design see Chowdhury et. al. (2019). For information on the MEPS MPC design, see RTI International (2019). A copy of the survey instruments used to collect the information on this file, are available on the [MEPS website](#).

2.0 Data File Information

This public use data set consists of two data files containing variables for linkage of the MEPS 2020 event-level data files. File 1, the H220IF1 or CLNK file, is used for linking the MEPS Conditions file with the MEPS event files; File 2, the H220IF2 or RXLK file, is used for linking the MEPS prescribed medicines event file with other MEPS event files.

The CLNK file 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. The RXLK file contains 6 variables and has a logical record length of 77 with an additional 2-byte carriage return/line feed at the end of each record.

2.1 Codebook Format

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

Identifier	Description
Name	Variable name
Description	Variable descriptor

Identifier	Description
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

2.2 Variable Naming and Source

In general, variable names reflect the content of the variable. All variables contained on Files 1 and 2 were derived from the CAPI.

2.3 Contents of File 1: Condition-Event Link File (CLNK)

File 1 (H220IF1) or the CLNK file, contains the variables needed to link each record on the MEPS 2020 Conditions file, HC-222, with one or more records on the MEPS 2020 event files, HC-220A, and HC-220D through HC-220H. Section 3.0 contains additional information on completing this linkage.

The ten-character variable DUPERSID uniquely identifies each person represented on the file. 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 file 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 2020 Conditions file, HC-222. The variable CONDIDX is the combination of the variables DUPERSID and CONDN (see HC-222 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 file 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 2020 event files, HC-220B through HC-220H. (EVNTIDX is not included on the 2020 Prescribed Medicines event file, HC-220A; rather, on this file the variable for linking with EVNTIDX on the CLNK file is LINKIDX.) There may be more than one record on the CLNK file 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-220G or
- OPAT - outpatient department visit event on MEPS release HC-220F or
- EROM - emergency room visit event on MEPS release HC-220E or

STAZ - inpatient hospital stay event on MEPS release HC-220D or

HVIS - home health visit event on MEPS release HC-220H

03 = PMED - prescribed medicine event on MEPS release HC-220A

CLNKIDX is the 29-digit number that uniquely identifies each record on the CLNK file and is the combination of CONDIDX + EVNTIDX. There is just one record on this file 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-220G

2 = OPAT - outpatient department visit event contained on MEPS release HC-220F

3 = EROM - emergency room visit event contained on MEPS release HC-220E

4 = STAZ - inpatient hospital stay event contained on MEPS release HC-220D

7 = HVIS - home health visit event contained on MEPS release HC-220H

8 = PMED - prescribed medicines event contained on MEPS release HC-220A

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 23, Panel 24, or Panel 25. Panel 23 is the panel that started in 2018, Panel 24 is the panel that started in 2019, and Panel 25 is the panel that started in 2020. The panel number is included as the first two digits of the DUID and DUPERSID.

2.4 Contents of File 2: Prescribed Medicines-Event Link File (RXLK)

File 2 (H220IF2) or the RXLK file, contains the variables needed to link some of the records on the MEPS 2020 Prescribed Medicines file, HC-220A, with one or more records on the MEPS 2020 files for other event types, HC-220B and HC-220D through HC-220G. Section 3.2 contains limitations on this linkage.

The ten-character variable DUPERSID uniquely identifies each person represented on the file. The variable DUPERSID is the combination of the variables DUID and PID. All ID variables begin with a 2-digit panel number. There may be more than one record on the RXLK file for a specific DUPERSID 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 2020 event files, HC-220B through HC-220G. There may be more than one record on the RXLK file for a specific EVNTIDX value. EVNTIDX has the 2-digit panel number added in the beginning, and a 2-digit event type number, 01, is added to the end.

LINKIDX is the 16-digit number that identifies the record(s) on the prescribed medicines file, HC-220A that link to an event record. There may be more than one record on the RXLK file for a specific LINKIDX value, and there may be more than one record on the HC-220A file for a specific LINKIDX value. LINKIDX has the 2-digit panel number added in the beginning, and a 2-digit event type number, 03, added to the end.

RXLKIDX is the 32-digit number that uniquely identifies each record on the RXLK file, and is the combination of EVNTIDX + LINKIDX. There is just one record on this file for each value of RXLKIDX, i.e., each unique combination of EVNTIDX + LINKIDX.

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-220G
- 2 = OPAT - outpatient department visit event contained on MEPS release HC-220F
- 3 = EROM - emergency room visit event contained on MEPS release HC-220E
- 4 = STAZ - inpatient hospital stay event contained on MEPS release HC-220D

For 1996-2004, records for purchases of insulin and diabetic supplies in a round were included in the Other Medical Expenses event files. Beginning with the 2005 file, these records are not included in the Other Medical Expenses file because the expenditures have always been included in the Prescribed Medicines file. As a consequence, there are no records in this file where the variable EVENTYPE = 6, the value used in 1996-2004 to identify OMED type of event record. Beginning in 2018, prescription medication is no longer asked for dental visit events. Therefore, there are no records in this file where the variable EVENTYPE = 5.

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 23, Panel 24, or Panel 25. Panel 23 is the panel that started in 2018, Panel 24 is the panel that started in 2019, and Panel 25 is the panel that started in 2020. The panel number is included as the first two digits of the DUID and DUPERSID.

2.5 ICD-10-CM, CCSR1X, CCSR2X, and CCSR3X

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 three CCSR categories (CCSR1X, CCSR2X, CCSR3X) using the v2020.3 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

These files are intended to be used in conjunction with other files. Specifically, the Conditions file (HC-222), the Prescribed Medicines Event file (HC-220A), and other event files (HC-220B through HC-220H).

3.1 Limitations/Caveats of the CLNK File

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

3.2 Limitations/Caveats of the RXLK File

The MEPS asked household respondents about prescribed medicines obtained from pharmacies, but not medications administered during a visit or other event. For many retail or mail order prescribed medicines, household respondents did not report that visits or other events were associated with the medicines. A medicine did not link to other event types when the respondent first reported the medicine in the prescribed medicines section of the survey, after the questions about other event types. (Sometimes, however, these medicines linked to other events in subsequent rounds.) Medicines that did not link to other event types have records in the prescribed medicines file but not in the RXLK file. For this reason, it would be difficult to use the MEPS to study prescribers.

When using the RXLK file, analysts should keep in mind that one event record can link to more than one prescribed medicine record. Conversely, a prescribed medicine record may link to more than one event record in the same event file and/or event records in more than one event file. When this occurs, it is up to the analyst to determine how to use the data.

3.3 National Health Interview Survey

Data from this file can be used alone or in conjunction with other files for different analytic purposes. Each MEPS panel can also be linked back to the previous years' National Health Interview Survey public use data files. For information on obtaining MEPS/NHIS link files please see the [MEPS website](#).

3.4 Using MEPS Data for Trend Analysis

First, of course, we note that there are uncertainties associated with 2020 data quality as discussed in the Survey Sample Information section of the Consolidated PUF document (HC-224). Evaluations described in that section suggest that care should be taken in the interpretation of estimates based on data collected in 2020 as well as in comparisons over time. Trend analyses are challenging since the advent of the COVID-19 pandemic resulted in uncertain data quality

for MEPS as well as standard benchmark sources such as the CPS, ACS, and NHIS while the pandemic also had an impact on the health and access to health care of the U.S. population. For such reasons, the extent to which 2020 health care parameters may differ from those of prior years is difficult to assess.

In terms of other factors to be aware of, MEPS began in 1996, and the utility of the survey for analyzing health care trends expands with each additional year of data; however, it is important to consider a variety of factors when examining trends over time using MEPS. The length of time being analyzed should also be considered. In particular, large shifts in survey estimates over short periods of time (e.g. from one year to the next) that are statistically significant should be interpreted with caution, unless they are attributable to known factors such as changes in public policy, economic conditions, or MEPS survey methodology.

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

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

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

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

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

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

3.5 Longitudinal Analysis

Panel-specific longitudinal files are available for downloading in the data section of the MEPS website. As has been done routinely in past years, the longitudinal file for Panel 24 comprises MEPS survey data obtained in Rounds 1 through 5 of the panel and can be used to analyze changes over a two-year period. Unlike past years for MEPS, in 2020 Panel 23 had data collected for a third year. As such, two-year and three-year longitudinal files will be developed for Panel 23. These can be used to analyze changes over the corresponding two-year or three-year period. Variables in the file pertaining to survey administration, demographics, employment, health status, disability days, quality of care, patient satisfaction, health insurance, and medical care use and expenditures were obtained from the MEPS full-year Consolidated files from the years covered by each panel.

For more details or to download the data files, please see Longitudinal Weight files at the [AHRO 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. Agency for Healthcare Research and Quality, Rockville, MD.

RTI International (2019). *Medical Provider Component (MEPS-MPC) Methodology Report 2017 Data Collection.* Rockville, MD. Agency for Healthcare Research and Quality.