Medical Expenditure Panel Survey—Medical Provider Component

2023 Plan for Matching MPC to HC Events

Deliverable OP3-26, V1.0

March 29, 2024

Prepared for

**Agency for Healthcare Research and Quality**

AHRQ, Center for Financing, Access & Cost Trends

5600 Fishers Lane

Rockville, MD 20857

Prepared by

RTI International

3040 Cornwallis Road

PO Box 12194

Research Triangle Park, NC 27709-2194

AHRQ Contract Number 75Q80121C00005

**This document was edited in December 2024 to redact potentially sensitive or proprietary information.**

Table of Contents

**Section Page**

[1. Introduction 5](#_Toc189128076)

[2. Preparing 2022 MPC Files for Editing 7](#_Toc189128077)

[2.1 Merging Blaise, Control File, and Frame Data 7](#_Toc189128078)

[3. Editing the 2022 MPC Medical File Data 9](#_Toc189128079)

[3.1 Basic Editing 9](#_Toc189128080)

[3.2 Managerial Edits Conducted During Data Collection 10](#_Toc189128081)

[3.3 Identification and Resolution of Duplicate Events 14](#_Toc189128082)

[3.4 Coding Text Responses 16](#_Toc189128083)

[3.5 SBD–OBD Overlap 16](#_Toc189128084)

[3.6 Process to Determine Whether an OBD Was an SBD During Editing 17](#_Toc189128085)

[3.7 Editing SBD Data 18](#_Toc189128086)

[3.7.1 SBD Encounter-Level Data Edits 18](#_Toc189128087)

[3.7.2 SBD Encounter-Level Final Disposition Code 19](#_Toc189128088)

[3.7.3 SBD Node-Level Data Edits 21](#_Toc189128089)

[3.8 Indicators for INDIAN\_HEALTH, CHARITY, MILITARY, or VA Type of Provider 24](#_Toc189128090)

[3.9 Mapping of ICD-10 Codes to CCSR Codes 26](#_Toc189128091)

[3.10 Data Provided by the Department of Veterans Affairs 26](#_Toc189128092)

[4. 2022 Medical File Editing Activities— Quality Control Checks 27](#_Toc189128093)

[4.1 Internal Checking by Primary Programmer 27](#_Toc189128094)

[4.2 Comparison of Data Distributions with Previous Years 27](#_Toc189128095)

[5. Summary of Key 2022 MPC Medical File Variables 29](#_Toc189128096)

[6. 2022 Medical File Pre-Matching Activities 35](#_Toc189128097)

[6.1 Obtaining the Household Survey Data 35](#_Toc189128098)

[6.2 Obtaining the MPC Data from the Edited Files 35](#_Toc189128099)

[6.3 Medical File Rollup Groups 36](#_Toc189128100)

[6.3.1 Definition of Rollup Groups 36](#_Toc189128101)

[6.3.2 Variables Created for Selected Rollup Groups 41](#_Toc189128102)

[6.3.3 After the rollup groups are constructed, selected variables are created that identify rollup group linkages for some of the groups. Compound Rollups 41](#_Toc189128103)

[6.4 Dealing with Rollup Groups During Matching 43](#_Toc189128104)

[6.5 Post-Matching Processing of Rollup Groups 43](#_Toc189128105)

[7. 2022 MPC Statistical Matching (Record Linkage) Process 46](#_Toc189128106)

[7.1 The Matching Algorithm 46](#_Toc189128107)

[7.2 Matching in Two Passes 48](#_Toc189128108)

[8. Variables Used in the 2022 Medical File Matching 50](#_Toc189128109)

[8.1 Blocking Variables 50](#_Toc189128110)

[8.2 Construction of Collapsed Contact Groups for Matching in 2022 51](#_Toc189128111)

[8.3 Matching Fields (Variables) 53](#_Toc189128112)

[9. 2022 Medical File Matching Process—Quality Control Checks 64](#_Toc189128113)

[10. 2022 MPC Home Health and Home Non-Health Data 66](#_Toc189128114)

[10.1 Preparing the HH and NHH Data for Editing 66](#_Toc189128115)

[10.2 Editing HH and NHH Data 66](#_Toc189128116)

[10.3 Matching of the HH Data 67](#_Toc189128117)

[10.4 Summary of Key 2022 MPC HH and NHH Variables 69](#_Toc189128118)

[11. 2022 MPC File Delivery and Preparing Codebooks 70](#_Toc189128119)

[12. Editing and Matching Procedures for the 2023 MPC 71](#_Toc189128120)

[12.1 Plans for Editing 71](#_Toc189128121)

Acronyms

| **Acronym** | **Definition** |
| --- | --- |
| AO | Administrative Office |
| AHRQ | Agency for Healthcare Research and Quality |
| BETOS | Berenson-Eggers Type of Service |
| CCSR | Clinical Classifications Software Refined |
| CPT-4 | Current Procedural Terminology—codes developed by the American Medical Association to describe medical, surgical, and diagnostic services |
| DCS | Data Collection Staff |
| DRG | Diagnosis-Related Group |
| ER | Emergency Room |
| FFS | Fee-for-service |
| GF | Global Fee |
| HC | Household Component |
| HCH | Home Care Health—same as HH |
| HCN | Home Care Non-Health—same as NHH |
| HH | Home Health—same as HCH |
| HHA | Home Health Agency |
| HOSP | Hospital component or event form |
| HS | Hospital Stay (i.e., inpatient) |
| ICD-10 | Internal Classification of Diseases, 10th edition |
| ID | Identifier |
| INST | Institution component or event form |
| LTC | Long-Term Care |
| MEPS | Medical Expenditure Panel Survey |
| MP | Medical Provider |
| MPC | Medical Provider Component |
| MR | Medical Records |
| MV | Medical Visit |
| NHH | Home Non-Health—same as HCN |
| NPI | National Provider Index |
| OBD | Office-Based Doctor |
| OP | Outpatient |
| OP1 | Option Period 1. MEPS MPC Contract Term. Also, referred to as 2021. |
| OP2 | Option Period 2. MEPS MPC Contract Term. Also, referred to as 2022. |
| PA | Patient Accounts |
| PHAR | Pharmacy component or event form |
| POC | Point of Contact |
| QC | Quality Control |
| RAPS | Radiologist, Anesthesiologist, Pathologist, Surgeon |
| SBD | Separately Billing Doctor |
| SOP | Sources of Payment |
| VA | Veterans Affairs |

# Introduction

The Medical Expenditure Panel Survey (MEPS) Household Component (HC) collects detailed medical information about each person in the household, including the use of medical services. During the survey interview, information about medical providers is collected along with authorization to contact these providers in the Medical Provider Component (MPC) of the MEPS. The HC also collects additional information about services provided and information on the cost.

The MPC portion of MEPS is particularly important to estimating the country’s healthcare costs because this component of MEPS collects information on the charges and payments associated with each medical event for a sample of people responding to the HC. Data from the MPC are more accurate since the data come directly from provider files. This information is collected from the provider to minimize the non-sampling error (e.g., recall error) that could exist if these data were collected solely from household respondents. Furthermore, for most medical events, a person might not know or pay attention to the total charges and payments associated with a medical service, particularly since many payments come from other sources such as Medicare, Medicaid, and private insurance companies.

After the MPC data are collected, the data are subjected to numerous edits and quality control (QC) checks before they are combined with the HC data. This stage of the processing helps to ensure the integrity and accuracy of the final estimates produced from MEPS. This process includes conducting basic edit checks and fill-ins, such as assigning missing codes to variables when appropriate, and cleaning the data to accurately reflect skip patterns in the instrument. Additionally, this process includes checking the distribution of key payment and charge variables. This check on the distributions is conducted to identify outliers and anomalous data, typically by comparing distributions against what was observed in previous MEPS MPC efforts. When outlier or anomalous data are detected, they are either edited programmatically or sent back to the data collection and abstraction staff for clarification.

To minimize the burden on providers, data are collected on all medical events for the survey reference year that are associated with each person in the HC subsample. A record matching process then occurs after the HC and MPC data are collected. The HC medical events for each patient are compared with complementary MPC medical events reported by the providers for the same patient. The goal is to correctly link provider data with data collected in the HC by patient and medical event.

Unique identifiers that would allow a direct merge of the HC and MPC files do not exist. For example, because of respondent recall error, the date of a particular medical event reported on the HC may differ slightly from the date the provider reported. Since unique identifiers of the reported events do not exist in a way that allows a direct linkage of HC and MPC records, a probabilistic matching approach is implemented. Several pre-matching steps are incorporated in the process to clean the data to facilitate matching. The likelihood of a match for each record is estimated, and a pair is classified as a match if the estimated likelihood exceeds a preset threshold.

This report serves two purposes. It summarizes the data editing and matching processes conducted for the 2022 MPC. It also discusses the plans for implementing these procedures for the 2023 MPC. With a few exceptions, the 20232 MPC will replicate the processes conducted for the 2022 MPC, to preserve comparability between years and to keep the editing and matching processes as efficient as possible. In the 2022 MPC (as in the 2021 MPC), when the only services received for an event were COVID-1 vaccines or test, any -1’s in the total charge were set to 0, and a flag was added to the file to indicate this edit. RTI will consult with AHRQ if we see similar situations in the 2023 MPC. Changes for the 2022 MPC-HC files and matching include a revision to the CCSRMATCH variables to expand the number of condition indicators used as matching variables and revisions to the arrays of BETOS variables, so that there is a one-to-one correspondence with the CPT codes.

In summary, input files from the data collection software system, control system, and various frame files are read, and an initial set of raw data files are constructed. The raw data are partitioned into the Medical File data (Hospital, Institution, office-based doctor [OBD], and separately billing doctor [SBD]) and Home Health/ Home Non-Health (HH/NHH) data files. This partitioning is done for efficiency purposes—similar data are kept together in the same file since they need to undergo similar phases of editing. For example, the Hospital, Institution, OBD, and SBD event forms ask for the same data (i.e., charges, payment, and reasons for discrepancy). Thus, data from all three sources are processed together so that their charge, payment, and reason-for-discrepancy data items are edited consistently and efficiently. After the data are edited, they are split back into their component parts, the MPC-HC matching phase occurs, and the final files are produced and sent to AHRQ for subsequent processing by other contractors.

# Preparing 2022 MPC Files for Editing

The following section discusses the process of preparing the data files for editing. Also included in this section is a summary of the identifier (ID) variables. Because of the hierarchy of the Medical Expenditure Panel Survey (MEPS) data, IDs are assigned at the contact group,[[1]](#footnote-1) provider, patient, event, and encounter (for separately billing doctors [SBDs]) levels.

## Merging Blaise, Control File, and Frame Data

The initial step in preparing the 2022 Medical Provider Component (MPC) files for editing was to combine various files from the data collection system. Data collection staff (DCS) used **Blaise** (the data collection system used for the MEPS MPC) during telephone contacts with respondents and to enter data from records after data abstraction. The output files from this system are known as the **Blaise files**, which include the following:

|  |  |
| --- | --- |
| * Hospital file * Office-based doctor (OBD) file * SBD file * Pharmacy file * Institution file | Home Health (HH) file, also referred to as the Home Care Health (HCH) file  Home Non-Health (NHH) file, also referred to as the Home Care Non-Health (HCN) file |

The 2022 cycle included files from the VA with cost and utilization data for pairs with VA providers (as was the case in the 2021 cycle). After the Blaise files and VA files were combined into one file, patient and provider demographic data were merged onto the concatenated data file. These demographic data consisted of patient data, including name, gender, and age (as of 6/1/2022), and provider data, including provider name and address from the sample frame files. The status (respondent, refusal, etc.) and the Edit Shop status variables were also merged onto the Blaise data. **Edit Shop** is the post-data collection process, started in 2011 for the 2010 MEPS MPC to identify data anomalies. Trained editors performed Edit Shop activities. When an edit was triggered by an event, an editor reviewed the event to determine if it was correct or if corrections were needed. If corrections were needed, the editor made these changes within Blaise so that the corrected data appeared on the raw data files prior to editing.

The Edit Shop variables that were merged onto the Blaise data included

* pair-level Edit Shop status
* number of times the event failed (i.e., number of times the edit was triggered by the event)
* most recent date the edit failed for the event (Edit Shop checks are run each time an event is revalidated in Blaise)
* most recent date the edit was checked for the event (Edit Shop checks are run each time an event is revalidated in Blaise)
* event-level status of the edit (1 = event failed edit; 8 = event passed edit; 9 = event failed edit, but the editor has reviewed and finalized)
* most recent date the editor reviewed the edit for the event (can be used to determine whether an event that failed the edit has been reviewed recently)
* editor’s comments for the most recent review (comments indicate the findings of the editor)

Having these Edit Shop variables on the raw and edited MPC files is useful when performing quality control (QC) activities. For example, if an inpatient event triggers Edit #7 (event cost is high or low for in-patient duration) and is reviewed by an editor, it may also appear as a data anomaly when charge data are examined during the QC process. Data editing staff can look at these variables and see that the event was reviewed in Edit Shop. Sometimes it is useful to review the editor’s comments, which explain why the event was or was not edited according to the flag triggered in Edit Shop.

Pairs of records sometimes exist on the raw Blaise files with the same MPC Person-by-Provider-by-Event ID (PMEID). In this situation, one record in the pair is coded as an **Original Abstraction**, and the second record in the pair was coded as a **Reabstraction**. This duplication occurs on purpose; for data capture QC purposes, a sample of the records is reabstracted and reentered, and the second entry is compared with the first. In these situations, the reabstracted record is kept on the data file and the originally abstracted record is dropped.

The Hospital, OBD, SBD, and Institution data are combined (i.e., concatenated) into one file. This file is called the initial **MPC Medical File** to distinguish it from the HH and NHH files.

# Editing the 2022 MPC Medical File Data

The following section summarizes editing Medical File data for the 2022 Medical Provider Component (MPC). The Medical File data include data from the Hospital, office-based doctor (OBD), Institution, and separately billing doctor (SBD) data collection forms.

Editing includes assigning consistency codes; performing a series of managerial edits for reviewing a variety of potential data issues, including identifying duplicate events; coding other text responses; identifying SBD–OBD overlap; determining whether an OBD was an SBD; performing additional edits for the SBD data; and coding provider-type indicators.

## Basic Editing

Numerous basic edits are performed to prepare the Medical File data for matching and imputation. These edits include assigning missing codes, specifically:

* Missing codes are assigned to correctly identify refusals and other types of nonresponse.
* The code for **Legitimate Skip** is assigned to some variables in cases where a missing response occurs because of skip logic in a stem question.
* Total payment and total charge are assigned the appropriate missing codes based on the codes recorded in the individual payment and charge fields. For example, “-5” indicates the charges are included in another event.

In addition to assigning appropriate missing codes, basic editing of the Medical File data includes recoding some variables to prepare them for matching and imputation. Examples of these recodes include the following:

* Resetting the global fee flag if any global fee data exist
* Recoding **other specify** strings for the sources of payment (SOP) variables into appropriate SOP codes
* Setting any missing reasons for discrepancy fields to **No** if one or more discrepancy items is Answered
* Recoding date variables when missing

Other variables are created throughout the editing process. Examples of these variables include

* Event type
* Berenson-Eggers Type of Service (BETOS) procedure codes
* Clinical Classifications Software Revised (CCSR) condition codes
* Flags for surgery, radiology, and laboratory tests

In preparation for matching, edits are performed to the condition and procedure codes within each event. These fields are cleaned by removing duplicates, separating multiple codes that occur in the same field, and ensuring that missing or blank codes appear at the end of the array of condition/procedure codes for an event. The condition and procedure codes are not ordered in any specific manner in the arrays, other than moving the codes so that any missing or blank codes are at the end of the array.

## Managerial Edits Conducted During Data Collection

A variety of data investigations are conducted throughout the MPC data collection. Data collection managers investigate and correct these issues (if necessary). These data edits generally look across events or pairs and are conducted in addition to the on-site Edit Shop run nightly during data collection. Spreadsheets that identify the specific events are prepared and shared with the data collection managers, who review the records and make necessary corrections to the data. **Exhibit 3.2** details the managerial edits that were conducted during the 2022 data collection.

Exhibit 3.2 2022 MPC Managerial Reviews

|  |  |
| --- | --- |
| Edit # | Edit |
| 1 | By patient, compare the distribution of events between the HC and MPC. Note that “difference” indicates that the number of HC events is greater than the number of MPC events. Cases are flagged if this difference exceeds a certain threshold, which varies by event type:   * If the number of HH events collected by MPC is less than half the number of HH events reported in HC, the case is flagged. * If HC and MPC IP/institution events have a difference of 1 or greater, the case is flagged. * If HC and MPC ER events have a difference of 2 or greater, the case is flagged. * If HC and MPC OP/MV events have a difference of 10 or greater, the case is flagged. |
| 2 | Identify OBDs with only S-type events for the same person and provider.  Starting in 2015, the location of service is changed to Physician Office when each event in the pair has at least one CPT that is 99201-99205, 99211-99215, or 99381-99387. |
| 3 | Identify and examine potential duplicate OBD and Hospital events that meet one of the following criteria:   * OBD, HS, OP, or ER events on identical or overlapping dates for the same person with the same total charge and/or total payment * IP hospital stays with overlapping dates   Starting in the 2020 cycle, this spreadsheet was expanded to include events with overlapping dates where the total OBD charges or payments was the same as a charge or payment for a Hospital event with the same set of dates.  Some potential duplicate events are resolved via a computer algorithm where the dates were the same and the total charge and/or total payments are the same. |
| 4 | Review OBD, Hospital (excluding IP), and SBD events with extreme values on payment and total charge variables, particularly large payment values from unexpected sources.  Starting in 2019, Home Health events with extreme values for total charge or payment were added to this review.  Excluded from review: Global fee, OBD, or OP events that include chemotherapy or cardiac care, OP lab events, and/or OP surgeries. |
| 5 | Examine events with missing reason for discrepancy when payment is less than charge and fee-for-service (FFS) events with missing charge or payment. |
| 6 | Examine extreme IP total charge and total payment values per stay day (abnormally large or small values); small values may be an indication that professional fees were collected. |
| 8 | Among Hospital events with the administrative office (AO) section complete, examine all information that was collected from AO about each SBD string as a QC step prior to SBD coding. |
| 9 | Examine other text responses for location of service. |
| 12 | Examine all events in pairs with a Hospital event with total charge greater than $75,000.  Starting in 2015, high-dollar events associated with CPT codes indicating preterm infants, acute kidney failure or end-stage renal disease, cancer, or heart conditions (except for only high blood pressure), where the ratio of total payment to total charge is between 0.20 and 0.90 are excluded from review. |

(continued)

Exhibit 3.2 2022 MPC Managerial Reviews (continued)

|  |  |
| --- | --- |
| Edit # | Edit |
| 13 | Examine instances where there is a mixture of FFS and capitated events from the same provider and for the same patient (ignoring pairs where payment was received from Medicaid for one or more events).  The following rules are used:   * Vaccines: If all events with only a vaccine as the CPT4 code or description are one type of payment routine and all other events (without any CPT4 or description of a vaccine) are the other type of payment routine, the event is left as is. * Charges included in another event (-5): If all events are capitated except for the event where charges are included in another event, the event is left as is. * Exams vs. Other Services (such as labs): If all examinations (as defined by CPT4 codes in the 99 series or descriptions) are one type of payment routine and all other events (without any CPT4 or description that includes an exam) are another type of payment routine, the events will be left as is. * Same date with different payment routines: If two events occur on the same date and one FFS and the other is capitated (CAP), events are considered correct if the examination codes or descriptions are always on one type of payment routine and never appear on the other type of payment routine. Note: Like rule 3 but both payment types show up on events from the same date. * Different locations: If all events with one type of payment routine are consistently at the same type of location while all events with the other type of payment routine are consistently at a different type of location, events are okay as is. All Hospital-based care (ER, OP, and IP) are considered the same type of location for purposes of this rule. * Switch of payment routine during the year: If the payment routine switches once during the year and there is at least 1 calendar month (as defined by month name, not days) between the switch, events are considered correct. * If none of the above rules apply and records are receipted, records are reviewed, and corrections are made as indicated by information in the records. * If none of the above rules apply and no records are receipted, the events are left as is. |
| 15 | Examine events with ADFROMER=“Yes,” but without a preceding ER event on the same or preceding day from the same provider. The variable ADFROMER indicates whether the patient was admitted from the ER or outpatient department. |
| 17 | Investigate OBD events with “HOSP” in provider name.  The following are excluded from review:   * Slash in provider name: If all events are associated with a provider where the word “hospital” appears before the slash (/) in the provider name, but the word hospital does not appear after the slash and the location of service for all events equals Physician’s Office, the pair is not flagged for review. * CPT codes and location of service: If the location of service for all events is Physician’s Office and the only CPT code is an evaluation and management code (99201, 99202, 99203, 99204, 99205, 99211, 99212, 99213, 99214, or 99215), the pair is not flagged for review. [Note: If there is more than one CPT code or if there is a description in addition to the CPT code, the pair is included in the spreadsheet and reviewed.]   All other events that do not meet the above criteria are included in the managerial review spreadsheet. Records are reviewed, when available. |

(continued)

Exhibit 3.2 2022 MPC Managerial Reviews (continued)

|  |  |
| --- | --- |
| **Edit #** | **Edit** |
| 20 | Identify providers who are not classified as MILITARY, VA, INDIAN\_HEALTH, and CHARITY with a large percentage of SOP responses that indicate they might be MILITARY, VA, INDIAN\_HEALTH, or CHARITY. This is used to assist in updating the master list of MILITARY, VA, INDIAN\_HEALTH, and CHARITY providers. |
| 22 | Compare counts of MPC events (observed) with HC events (expected) for all sample pairs. |
| 25 | Examine events that have duplicate OBD and Hospital PMEIDs. |
| 26 | Examine Pharmacy events collected with $0 patient payment, third-party payer = 21 and $0 third-party amount. |
| 28 | Review capitated event that occurred outside California.  The following situations are excluded from review:   * L5: If all events are associated with an L5 provider, the events are left as is. * Same charge for office visits: If there is more than one event and all are simple office visits with the same charge amount (must be a value—not don’t know or refused), the events are left as is. * Vaccines: If all events with ONLY a vaccine as the CPT4 code or description are one type of payment routine and all other events (without any CPT4 or description of a vaccine) are the other type of payment routine, the events are left as is. * Charges included in another event (-5): If all events are capitated except for the event where charges are included in another event, the events are left as is. * Exams vs. other services (such as labs): If all examinations (as defined by CPT4 codes in the 99 series or descriptions) are one type of payment routine and all other events (without any CPT4 or description that includes an exam) are another type of payment routine, events are left as is. * Same date with different payment routines: If two events occur on the same date and one is FFS and the other is CAP, events are left as is if the examination codes or descriptions are always on one type of payment routine and never appear on the other type of payment routine. Note: This is like rule 3 but both payment types show up on events from same date. * Different locations: If all events with one type of payment routine are consistently at the same type of location while all events with the other type of payment routine are consistently at a different type of location, events are left as is. Please note that all hospital-based care (ER, OP, and IP) should be considered the same type of location for purposes of this rule. * Switch of payment routine during the year: If the payment routine switches once during the year and there is at least 1 calendar month (as defined by month name, not days) between the switch, events are left as is. * If none of the above rules apply the events are included in the spreadsheet. Records are reviewed when available. |
| 29 | Examine Pharmacy events where the patient is age 40 or younger and Medicare is a payer.  Excluded are events for people where Medicare is a payor on the OBD or Hospital file, and events for people with a condition (ICD code) of ALS or chronic kidney disease on the OBD or Hospital file. |
| 30 | Review Pharmacy events with more than one event on the same day for the same patient with the same NDC code. |
| 31 | Review Pharmacy events with large amounts in the patient payment variables or third-party payer amount the same as the MPC year, the year before, or the year after. |
| 34 | Examine events with 4 or more sources of payment fields with positive payments where one or more of the payments (excluding patient payment) is exactly $2.00. This review also examines other events where any payment field (excluding patient payment) is exactly $2.00. |
| 35 | Review Hospital events where at least one CPT4 code has ‘26’ as a modifier (which indicates the profession component of a service or procedure). |

(continued)

Exhibit 3.2 2022 MPC Managerial Reviews (continued)

|  |  |
| --- | --- |
| **Edit #** | **Edit** |
| **Additional Managerial Review Edits for SBD** | |
| 50 | Examine SBD/OBD overlap identified after SBDs are collected. This is a listing of completed OBDs and SBDs with the same charge or payment for the same patient and same hospital event. |
| 60 | Examine SBD/Hospital potential duplicates by looking for events on the same day with SBD and Hospital charges the same and/or SBD and Hospital payments the same (excluding zero payments if that is all that is the same). |
| 70 | Examine SBD potential duplicates—responding SBD nodes or encounters for the same patient with the same Hospital ID, (nonzero) payment or charge, date, role, and specialty. |

ALS = amyotrophic lateral sclerosis; AO = administrative office; CAP = capitated; CPT = Current Procedural Terminology; DCS = Data Collection Staff; ER = emergency room; FFS = fee-for-service; HH = Home Health; HOSP = hospital; ICD = International Classification of Diseases; ID = identifier; IP = inpatient; HC = Household Component; MPC = Medical Provider Component; NDC = National Drug Code; OBD = office-based doctor; OP = outpatient; QC = quality control; SBD = separately billing doctor; VA = Veterans Affairs.

Note: Edit numbers 7, 10, 11, 14, 16, 18, 19, 21, 23, 24, 27, 32, and 33 are not used.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

## Identification and Resolution of Duplicate Events

Duplicate records are created when data are entered two or more times for the same person, provider, and medical event. Duplication in data collection can occur for several reasons. For example, sometimes two or more Household Component (HC) provider identifiers (IDs) that point to the same provider are unable to be connected before data collection, and sometimes the HC field interviewer inadvertently keys the same event twice for a person–provider pair.

Sometimes the data (including charges, payments, location, services provided, diagnosis, etc.) are the same for the duplicate events, and other times the data vary in small ways between the two events. Duplicated data may be collected within the same provider or between different providers. The objective is to identify these duplicate events, and then review and edit as needed to produce accurate, useful data for the MPC-HC matching process.

Potential duplicate events are identified and examined. These potential duplicate events meet one of the following criteria:

* OBD, outpatient (OP), inpatient, emergency room (ER), or Institution events on identical or overlapping dates for the same person with the same total charge and/or total payment
* Inpatient hospital stays (HSs) with overlapping dates

As in prior cycles of the MPC, data collection managers will review and resolve most potential duplicate events for the 2023 MPC. Managerial review of duplicate events often results in merges of either data collection contact groups or person–provider pairs. This information is used in matching: providers associated with the merged data collection contact group or merged person–provider pairs are placed into the same provider block group.

The process for identifying and resolving these duplicate situations has evolved each year based on lessons learned. The current process for identifying duplicates classifies them into four groups (Type 1 – Type 4) as defined below.

* To be classified as “Type 1” or “Type 2” duplicates, the dates of service, and the value of FEEORCAP must be the same. In addition, total charges must be the same and at least one of the events must have a nonmissing payment. If both payments are nonmissing, then the payments also need to be the same. The only difference between “Type 1” and “Type 2” duplicates is that “Type 1” duplicate events have the same provider ID and “Type 2” duplicate events have different provider IDs.
* To be classified as “Type 3” duplicates, the dates of service must be overlapping, and values for provider ID, FEEORCAP, CPT codes and SBDs must be the same. Hospital events are flagged as “Type 3” only if the location of service is the same or location of service is OP or somewhere else. There are no location-of-service stipulations for OBD events. Events are excluded from being classified as “Type 3” duplicates if they have any of the following characteristics:
  + ER and OP on the same day
  + OP events within inpatient stays
  + consecutive hospital stays
  + multiple events on the same day but different values for FFS/Capitated
  + hospital inpatient stays that overlap by one or more days
  + events with different values for both payments and charges
* To be classified as “Type 4” duplicates, the dates of service must be overlapping, and the events must not be classified as any of the other three types of duplicates. Values for payments and/or charges must be equal.

Starting in the 2015 MPC, at AHRQ’s request, some of the duplicates that are classified as “Type 1” and “Type 2” duplicates were resolved programmatically instead of through manual review to reduce the amount of manual review. The following procedure was used. OBD events (including the OBD S-type events that are to be moved to the SBD file) were classified as having medical visit (MV) as location of service. The priority for the event to keep was based on location of service (HS, then ER, then OP, and finally MV). For example, if one of the events had location of service HS and the other had location of service MV, then the event with location of service HS was flagged as the master event. If the location of service was the same for both events (e.g., both HS or both MV), then the one with the highest PMEID was kept as the master event (the event with the larger PMEID would have generally been collected later in the data collection period). SBD events were excluded, and OBD or Hospital events where either both events contained an SBD or only one of the events contained an SBD were left on the spreadsheet for review by data collection managers. In 2022, 64 duplicate sets that were classified as “Type 1” duplicates and 321 duplicate sets that were classified as “Type 2” duplicates were resolved programmatically.

Starting in the 2016 MPC, at AHRQ’s request, some of the duplicates that were classified as “Type 3” and “Type 4” duplicates were resolved programmatically instead of through manual review to reduce the amount of manual review.

RTI continued manual review for the following “Type 3” and “Type 4” duplicate situations:

* “Type 3” and “Type 4” sets with one or more inpatient events.
* Sets with more than two events.
* Sets where total charges or total payments are not identical.
* Sets that include a global fee.
* Sets where the location of service is identical for both events.

When both records have SBDs that are different, or only one record has SBDs and that record would not be chosen as the master, then the events are manually reviewed.

The remaining duplicate sets are resolved programmatically by auto coding these duplicate events like the procedure used for the “Type 1” and “Type 2” duplicates (events with the same dates within or across providers). Specifically, if payments were missing for one event and non-missing for the other, then the non-missing payment is moved into the payment field for the other event. Then, the master event is chosen using the priority ordering that is used for the “Type 1” and “Type 2” duplicates for picking the master event (priority ordering is based on location of service, in the order of HS, ER, OP, and MV). In 2022, one duplicate set that was classified as a “Type 3” duplicate and 25 duplicate sets that were classified as “Type 4” duplicates were resolved programmatically.

## Coding Text Responses

All text responses are reviewed, evaluated, and assigned appropriate codes so that they can be used during analyses and processing. The text responses come from survey questions where the responses differ from the options listed in the survey. The DCS enters the respondent’s description for any “other” responses.

If the Location of Service text response does not exactly match the definition of an existing code (ignoring spelling and spacing discrepancies), then the text response is compared with previous years’ responses and coded accordingly. This check is performed to preserve comparability between MEPS years. If the text cannot reasonably be assigned to any of the codes, the response is coded 96 (Something Else). Internet searches are also used to help determine if one of the existing codes (other than 96) might be appropriate.

## SBD–OBD Overlap

Question B3 of the OBD form asks OBD points of contact whether the patient received services in one of these locations:

|  |  |
| --- | --- |
| * Physician’s office * Hospital as an inpatient * Hospital OP department | Hospital ER  Somewhere else |

Responses “hospital as an inpatient,” “hospital outpatient department,” and “hospital emergency room” are called **S-type OBD events** because these are SBD charges for events that occur in a Hospital. If that Hospital responded to the MPC, one would expect some mention of the OBD event noted in the SBD portion of the Hospital event. To maintain the integrity of a medical event as defined by the MPC, all OBD S-type events are removed from processing, and if possible, these events are linked to the SBD mentions of HS. OBD S-type events that cannot be linked to any SBD mention are discarded from the analysis (i.e., these are not eligible as stand-alone events for the MPC-HC statistical matching process). No further review of S-type events that cannot be linked to a Hospital event is conducted. However, as a part of edit # 2 listed in **Exhibit 3.2**, all OBDs with only S-type events for the same person and provider are reviewed.

To minimize data collection costs and reduce respondent burden, RTI attempts to link OBD S-type events to SBD mentions before SBD data collection. If a probable match is found, then the SBD is not contacted, and the data collected on the OBD form are transferred to the SBD form as part of the data editing discussed in this report. Potential matches between the OBD and SBD samples for the MPC were identified using fuzzy matching text comparisons.

RTI also attempts to link S-type OBD and SBD data after SBD data collection for two reasons: (1) to catch any SBD–OBD overlaps missed in the pre-data collection matching phase, and (2) to select the record in the matched pair that contains the more complete data. In addition to contact information, payment and charges are also considered in this process. All OBD events are considered in this editing step, not just those classified as S-type events. This step is conducted to capture those OBD events that have an incorrect location of service classification.

The final version of the 2022 MPC SBD encounter-level file consists of 9,591 unique records (encounters). Of these, 5.3 percent of the encounters were collected in the OBD sample and did not overlap an SBD encounter. For the 2021 MPC, the comparable percentage was 7.1 percent.

## Process to Determine Whether an OBD Was an SBD During Editing

The rules that are used to identify an OBD that is an SBD differ according to whether the OBD is an S-type event. An OBD is called an S-type event if the location of service (variable RCSRV from the OBD event form) is either hospital as an inpatient, hospital in the OP department, or the hospital ER.

For OBD S-type events, the date of the OBD is considered in relation to dates of Hospital events occurring concurrently or close to the OBD date. For each S-type OBD event, the date of the OBD event is compared with dates of Hospital events that are close to the OBD event date. The logic for determining if an S-type OBD event is an SBD for a Hospital event is as follows:

* If the OBD event date occurs sometimes within an inpatient HS, including the first and last days of the stay, it is considered an SBD for the Hospital event.
* Otherwise, if the OBD event date occurs 1 day before or 1 day after an inpatient HS, it is considered an SBD for the Hospital event.
* Otherwise, if the OBD event date occurs on the first day of an ER/OP Hospital event, the OBD event is considered an SBD for the Hospital event.
* Otherwise, if the OBD event date occurs 1 day before or 1 day after an ER/OP hospital event, the OBD event is considered an SBD for the Hospital event.

Because an S-type OBD event can be linked to more than one Hospital event, a priority is assigned to the hospital event to which the S-type OBD event is linked above by the numbering (1 to 4). The priority means that an S-type OBD event is linked to an inpatient stay if possible; otherwise, the S-type OBD event is attempted to be linked to an ER/OP event. Since 2010, all S-type OBD events are considered ineligible for matching regardless of whether they can be linked to a Hospital event.

Non–S-type OBD events (i.e., medical visit events) are also examined and classified as SBDs if the date of the OBD event and dates of Hospital events are close or concurrent. Specifically, if the OBD event is strictly within an inpatient HS (i.e., it does not equal the first day or the last day, but equals one of the days in between), then the OBD is an SBD for the event.

## Editing SBD Data

Numerous edits are performed to prepare the SBD data file for delivery. The SBD file contains data pertaining to Hospital events where a physician billed separately from the hospital, for services provided during the hospital event. For example, a surgeon, anesthesiologist, or primary care physician could see a patient more than one time during a hospital event, and each of these doctors may bill separately for their services.

The hospital event by an SBD combination (e.g., physician) is referred to as an **SBD node** in MEPS. As noted above, each SBD node could have several visits during an inpatient stay, and each visit may be charged separately. Each visit provided by an SBD node is known as an **SBD encounter** in the MEPS MPC.

Edits are first performed to the encounter-level file (**SBDENCTR** is the name of the encounter-level file). The encounter-level file is then summarized to the node level, and additional edits are performed (**SBD\_SUM2** is the name of the node-level file). Edits at both the encounter and node levels are discussed in this section.

### SBD Encounter-Level Data Edits

Data collection managers perform the following types of reviews (listed in **Exhibit 3.2**) for the SBD data. All instances are reviewed and corrected when necessary.

* Examine events where SBD total charge or payment is the same as the Hospital charge or payment. When the SBD and Hospital amounts are the same, either the SBD charges are included in the hospital bill or a data collection abstraction error of some sort has occurred. The data collection managers review the SBD and hospital events and make the appropriate corrections.
* Examine SBD encounters on the same day for the same patient with the same role and specialty and same charge or payment to see if these are duplicate encounters, or if the procedure or event can be billed multiple times a day.
* Examine OBD and SBD events for the same patient associated with the same HS to determine if duplication has taken place among the OBD and SBD events.

The data are also reviewed to ensure that certain variables are consistent across encounters in the same node. The variables considered in this review include the following:

* SBD variables
* Role of SBD
* Class of SBD (whether the doctor performed any of the following types of services: radiology, anesthesiology, pathology, surgery, none, or don’t know)
* Specialty variables
* Indicator of whether the SBD events are complete, partially complete, partially complete with critical items missing, or not complete

These variables are generally consistent. However, when inconsistency occurs, the value from the first encounter in the node is used.

Several variables unique to the SBD encounter-level data are produced. These include BUNDLTYP, which identifies the type of global fee bundle, as well as several variables carried over from the corresponding Hospital event.

Other variables created for the encounter-level file include the Final Disposition Codes and the Payment Amount and Source of Payment variables.

### SBD Encounter-Level Final Disposition Code

The final disposition of each SBD encounter (stored in the variable FIN\_DISP) is coded based on whether the encounter is considered complete and on the disposition of the node, encounter, pair, or group at the end of SBD data collection. **Exhibit 3.3** shows the values for the node-level FIN\_DISP status code variable and the corresponding values of final disposition of the node. Some status codes that are assigned at the pair and group levels should logically carry over to the encounter level (e.g., refusals and disavowals). Within each node, encounters almost always have the same disposition status code; when encounters differ, the disposition of the primary node is used.

Exhibit 3.3 Mapping of the Final Disposition of the Node to FIN\_DISP—2022 MPC

|  |  |
| --- | --- |
| FIN\_DISP Label | Node Status Label |
| Partially Complete | Partially Complete |
| Complete | Complete |
| Other Refusal | Other Refusal |
| Unlocatable | Unlocatable |
| Unlocatable | Insufficient info to locate |
| No data, info purged | No data, info purged |
| Refuses to retrieve | Refuses to retrieve |
| Included in Another Nonresponse OBD | Included in Another Nonresponse OBD |
| Disavowal 1 | Disavowal 1 |
| Disavowal 2 | Disavowal 2 |
| Referred / Copied Physician | Referred / Copied Physician |
| Dept Head Follow-up | Dept Head follow-up |
| Not an SBD | Not an SBD |
| Not an SBD | Final — Not SBD per AO |
| Other | Empty Node or Pending Code |
| Other | Reached Max Call Attempts |
| Other | Contact Group Not Worked |
| Other | Other Nonresponse |
| Other | Abstracted but not Fielded — HOSP/INST EF created in error |
| Other | SBD is a duplicate of a non-SBD |
| Other | Abstracted but not Fielded — SBD Node Created from HOSP Event Form |
| Other | Coding incomplete |
| Abstracted in Error | Abstracted in Error |
| Abstracted in Error | Duplicate of existing node from same hospital event |
| Included in Another SBD | Included in Another SBD |
| Included in Hospital | Included in Hospital Event |
| Included in OBD | Included in OBD |

(continued)

Exhibit 3.3 Mapping of the Final Disposition of the Node to FIN\_DISP—2022 MPC (continued)

|  |  |
| --- | --- |
| FIN\_DISP Label | Node Status Label |
| Held nodes (subsampled and not fielded) | Abstracted but not Fielded — SBD Node built into Sample |
| Node is a leaf | Node is a leaf |

AO = administrative office; EF = Event Form; HOSP = hospital; INST = institution; OBD = office-based doctor; SBD = separately billing doctor.

The node-level file that RTI delivers includes all nodes that are linked to complete or partial Hospital events. This includes some nodes that are not fielded (e.g., because they are determined to not be SBDs prior to data collection, because there was not enough information to attempt to locate the SBD, or because the node was held from fielding due to subsampling).

Complete or partially complete SBD encounters can include SBDs collected on the SBD event form, OBD events (including S-type events) that are reclassified as SBDs, and OBDs that are determined to overlap (or duplicate) an SBD and whose data are copied into the SBD fields.

### SBD Node-Level Data Edits

The SBD node-level file contains the following:

* SBD nodes and encounters that are summarized to the node level that are collected on the SBD event form
* OBD events that are classified as SBD events because of the location of service and date in relation to the HS, and OBD medical visit events that are classified as SBD events based on the date in relation to the Hospital event
* OBD events that are determined to be the same as an SBD event during the SBD–OBD overlap review
* OBD events with the same Hospital event ID and SBD provider ID as an SBD encounter are placed in the same node as the SBD encounter for processing.

A key step in editing the node-level SBD data is to summarize the encounter-level data to the node level. Prior to summarizing to the node level, SBD Source of Payment and Payment Amount variables are created at the encounter level. These variables are created for the following payment sources:

|  |  |  |
| --- | --- | --- |
| * Patient or patient’s family * Medicare * Medicaid * Private insurance | * VA/CHAMPVA * TRICARE * Other Federal payment source | * Other State/local payment source * Worker’s compensation * Other payment source |

Collapsing the encounter-level source of payment and payment amount variables to the node level is not as easy as adding payments. The process is complicated because of the varying values assigned to the payment variables across encounters within the same node, such as codes for missing payment sources and amounts, as well as inconsistencies in the data between source of payment and amounts of payment by sources. For example, the source of payment may indicate that private insurance made a payment, but the payment amount for private insurance is missing or vice versa. To summarize these payment variables to the node level so that the node-level values are indicative of all encounters in the node, the SBD data are processed by performing edits provided by AHRQ and the HC contractor who will use these data in the subsequent MEPS estimation phase. The end results of these edits are node-level variables for the SBD source of payment (SBDSOP1-10) and SBD payment amount (SBDAMT1-10). These edits prepare the payment source and amount variables for the contractor’s imputation process.

Final disposition codes are also created at the node level. In general, all encounters within each node have the same disposition code except for occasional nodes that contain encounters with disposition codes of complete and partially complete. The final disposition of these nodes is assigned to “complete.”

Several variables unique to the node-level data are produced. A variable is created for the number of SBD nodes in a global fee bundle and a flag for out-of-scope SBD nodes . After the MPC-HC Hospital event matching, two additional variables related to the matching results are created—(1) an indicator of whether the Hospital event is matched to an HC event and (2) a variable to flag out-of-scope nodes that are matched to an HC event .

The following Hospital variables are merged onto the node-level data: Hospital IDs, Hospital event setting, Hospital total charge, and flags for radiology, surgery, and lab.

Finally, several node-level pre-edits are performed that AHRQ and the HC contractor provided to us.

**Exhibit 3.4** displays the average number of SBD encounters per responding node by event type and several Hospital-level variables for the most recent 3-year period. Generally, the yearly averages are similar, and the patterns for the average number of SBD encounters across Hospital event types and Hospital event characteristics follow an expected pattern. For example, as charges and payments associated with the HS increase, the number of SBD encounters per node increases. In 2022, the average number of SBD encounters per node for events at institutions was 4.4, compared to 3.0 in 2021. When examining the 2022 total charge for Hospital events, the encounters per node range from an average of 1.0 SBD encounter per node to 1.4 SBD encounters per node as charges increase from $0 to over $10,000.

Exhibit 3.4 Average Number of SBD Encounters per Responding Node, 2020–2022

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Category | 2020 | 2021 | 2022 |
| Total Sample | Total | 1.2 | 1.2 | 1.2 |
| Hospital Event Type | Inpatient | 1.7 | 1.6 | 1.5 |
| Emergency Room | 1.0 | 1.0 | 1.0 |
| Outpatient | 1.0 | 1.0 | 1.0 |
| Institution | 4.3 | 3.0 | 4.4 |
| Total Charges on Hospital Event | Missing | 3.0 | 1.4 | 1.0 |
| Zero | 1.0 | 1.0 | 1.0 |
| $1–$250 | 1.0 | 1.0 | 1.0 |
| $251–$1,000 | 1.0 | 1.0 | 1.0 |
| $1,001–$10,000 | 1.0 | 1.0 | 1.0 |
| > $10,000 | 1.5 | 1.5 | 1.4 |
| Total Payments on Hospital Event | Missing | 1.2 | 1.3 | 1.0 |
| Zero | 1.1 | 1.1 | 1.0 |
| $1–$250 | 1.0 | 1.0 | 1.0 |
| $251–$1,000 | 1.0 | 1.0 | 1.0 |
| $1,001–$10,000 | 1.2 | 1.2 | 1.2 |
| > $10,000 | 1.8 | 1.6 | 1.6 |
| Data Collection Mode for Hospital Event | Telephone | 1.9 | 2.7 | 2.7 |
| Hard copy | 1.2 | 1.2 | 1.1 |

SBD = separately billing doctor.

Note: The statistics are based on unique (nonreplicated) Medical Provider Component events.

Source: 2020-2022 Medical Expenditure Panel Survey – Medical Provider Component.

**Exhibit 3.5** displays the average number of nodes per Hospital event (excluding Hospital events with no nodes) for the most recent 3-year period. The average number of nodes per Hospital event in 2022 (2.2) is consistent with the averages in 2021 (2.3) and 2020 (2.3). The averages follow logical patterns. For example, the 2022 average number of nodes per Hospital event is 5.1 for Hospital events with charges > $10,000 and 1.4 for Hospital events with charges of $1–$250. The higher charges and larger number of nodes per Hospital event are likely the result of a greater number of stay nights.

Exhibit 3.5 Average Number of Nodes per Hospital Event (Excludes Hospital Events with No Nodes), 2020–2022

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group | Category | 2020 | 2021 | 2022 |
| Total Sample | Total | 2.3 | 2.3 | 2.2 |
| Hospital Event Type | Inpatient | 7.7 | 7.5 | 7.1 |
| Emergency Room | 3.1 | 3.1 | 2.8 |
| Outpatient | 1.8 | 1.8 | 1.8 |
| Institution | 1.7 | 1.8 | 2.8 |
| Total Charges on Hospital Event | Missing | 2.9 | 2.9 | 1.9 |
| Zero | 3.5 | 3.5 | 3.2 |
| $1–$250 | 1.5 | 1.5 | 1.4 |
| $251–$1,000 | 1.7 | 1.7 | 1.6 |
| $1,001–$10,000 | 2.4 | 2.5 | 2.2 |
| > $10,000 | 5.9 | 5.6 | 5.1 |
| Total Payments on Hospital Event | Missing | 2.1 | 2.6 | 1.8 |
| Zero | 2.6 | 2.8 | 2.4 |
| $1–$250 | 1.7 | 1.7 | 1.6 |
| $251–$1,000 | 2.2 | 2.2 | 2.1 |
| $1,001–$10,000 | 3.6 | 3.7 | 3.4 |
| > $10,000 | 7.8 | 7.5 | 6.8 |
| Data Collection Mode for Hospital Event | Telephone | 1.3 | 1.5 | 2.0 |
| Hard copy | 2.4 | 2.3 | 2.2 |

Note: The statistics are based on unique (nonreplicated) Medical Provider Component events.

Source: 2020-2022 Medical Expenditure Panel Survey – Medical Provider Component.

## Indicators for INDIAN\_HEALTH, CHARITY, MILITARY, or VA Type of Provider

Four indicator variables that denote whether an MPC provider is likely to be associated with the Indian Health Service, provide charity care, be a provider for the U.S. military, or be a provider with the Department of Veterans Affairs (VA) have been created, starting in 2011. These four variables are INDIAN\_HEALTH, CHARITY, MILIARY, and VA and take values 0 or 1 for each event on the file. These variables only indicate that the provider may be associated with that payment source, and not that the MPC event had that source of payment.

The creation of these indicators started with the construction of a master list of names and addresses of each type of provider. The lists for all the provider types used for the 2022 MPC were based on the following sources:

##### Indian Health

Indian Health Service Facilities Directory (downloaded January 24, 2023 from <https://www.ihs.gov/findhealthcare/>).

##### Charity

Directory of Federally Qualified Health Centers and Look-Alikes from Health Resources and Services Administration (HRSA) (downloaded January 24, 2023 from https://data.hrsa.gov/data/download). The specific file used is “Health Center Service Delivery and Look-Alike Sites.”

Charity clinics from the National Association of Free and Charitable Clinics (downloaded May 26, 2020 from https://www.nafcclinics.org/find-clinic).

##### Military

TRICARE medical treatment facilities (downloaded January 24, 2023 from https://www.tricare.mil/FindDoctor/mtf?country=-18&pageNo=1&pageSize=5&view=map).

##### VA

VA facilities (downloaded January 24, 2023) from https://www.va.gov/directory/guide/rpt\_fac\_list.cfm

Providers that were determined to be one of the four types in a prior MPC cycle were kept on the list. These lists are matched to the MPC sample providers to determine whether each sample provider can be classified into any of the four categories. Matching on telephone number and text matching of names and addresses is used to initially classify MPC providers for each indicator, followed by a manual review. Searches in the provider name for strings such as “VA” (where the State is not Virginia), “Veteran,” “CBOC” (Community-based Outpatient Clinic), and similar text strings are used to identify additional VA facilities among the MPC providers.

The MPC data are also used as a source for identifying these types of providers. Proportions of the events for each provider that reported SOP as Indian Health, sliding scale or charity, military, and VA are computed. If more than half of the events for a provider have one of these SOP checked, but the provider is not classified as such, then the provider’s name and address are printed and examined using Internet searches for possible addition to the master list and reclassification. Sensitivity analysis has not been done on the 50 percent threshold. However, most of the providers flagged by this rule are not Indian Health, charity, military, or VA and the level of effort required for the review is low.

Beginning with the 2020 MPC, RTI examined providers that are not classified as one of the four types and also contain “County” or “District” in the provider’s name to determine if any should be reclassified as CHARITY. Similarly, providers with these text strings in the provider’s name will be examined for the 2023 MPC and coded as “CHARITY” if this is appropriate. In addition, beginning with the 2022 MPC, RTI used the hospital ownership field of the Centers for Medicare and Medicaid Services Hospital General Information file (<https://data.cms.gov/provider-data/dataset/xubh-q36u>) as a preliminary indicator of government ownership.

If a provider is identified as one of the four types, then it is added to a historical list that is maintained from year to year.

The master lists of the four types of providers were last updated for the 2022 MPC . For the 2022 MPC, RTI suggests appending the current military lists with the new names and addresses of the nine military bases that were renamed during 2023.

## Mapping of ICD-10 Codes to CCSR Codes

RTI maps the ICD-10 codes into CCSR codes, using the mapping from the Healthcare Cost and Utilization Project (HCUP) website, and these CCSR codes are then mapped into CCSRMATCH code for use in the matching. The December 2022 (V2023.1) version of the mapping was used for the 2022 MPC files. Sometimes partial ICD-10 codes are collected in the MPC. To be considered a “partial” ICD-10 code, the code must be a minimum of 3 characters and/or digits and match the beginning characters of a complete ICD-10 code. During the 2019 MPC, RTI and AHRQ worked together to develop rules and an algorithm for mapping partial ICD-10 codes into CCSR codes.

## Data Provided by the Department of Veterans Affairs

Pharmacy and Hospital pairs in the sample where the provider was identified as a Veterans Affairs (VA) provider were sent to AHRQ who requested the charges and utilization data for the pairs from the VA. RTI incorporated the VA events with the data for the non-VA Hospital and Pharmacy pairs that were collected in Blaise.

Prior to incorporating the VA data with the MPC data from Blaise, RTI and AHRQ examined the Hospital and Pharmacy data for duplicate events or other anomalies such as zero or negative quantities. In total, the VA provided 10,816 Pharmacy events of which 10,802 were retained and combined with the MPC-collected Pharmacy data, and 7,379Hospital events of which 6,916 were retained and combined with the MPC-collected Hospital data.

##### NonPharmacy (Hospital) Edits

VA Hospital variable were renamed or recoded to be consistent with the names and values from the MPC Blaise data. MPC Hospital variables that were not included in the data from the VA were logically assigned consistency codes or other values for VA Hospital events.

The VA Hospital file included the location of service. Additional categories on the VA Hospital file identified various types of telehealth, Home Health, and long-term care encounters.

The VA Hospital data did not include charge variables, and all MPC charge variables were set to “don’t know” (-8) for the Hospital events provided by the VA. The VA provided the payment variable VAPRODCOST, which is the VA third-party payment amount; the value of this variable was assigned to the MPC variable VAPAYM. All other MPC 3rd party PAYM fields were set to $0 when the VA data was combined with the MPC data. The VA data contained a variable which indicated whether there was any patient copay, but this patient copayment amount was not provided. The patient copay amount variable PATPAYM was set to $0 when there was no patient copay, and was set to missing when there was an indication that there was a patient copay. As a result, all of the events from the VA have a missing value of TOTLCHRG, and many are missing TOTLPAYM (when there was a patient copayment amount but it was missing). Any VA event with a VAPAYM amount were considered eligible for matching (unless they were ineligible for some other reason, such as being in a rollup group or having a nonpositive weight); usually, MPC events missing both total charge and total payment are not eligible for matching.

# 2022 Medical File Editing Activities— Quality Control Checks

As part of the effort to produce useful and high-quality data, various checks are implemented throughout the Medical Provider Component file pre-matching and matching procedures. The approach to quality control includes internal checking by the programmer with primary responsibility for the activity and comparison of data distributions with previous years. Each of these activities is described below.

## Internal Checking by Primary Programmer

The primary programmers perform internal checks concurrent with running the editing and matching programs. These internal checks include, but are not limited to, case reviews, coding checks, and consistency checks. When a new variable is created, a cross tabulation is run to check the programming code that created the variable. The primary programmers also examine extreme values of continuous variables to identify outliers.

Primary programmers also conduct quality checks after each major step of their activities. These checks include the following:

* Tabulating some key variables such as event type, conditions, and procedures. These tabulations help give an understanding of the data and help identify any issues potentially affecting the quality of the matching results.
* Manually reviewing a sample of duplicated records, consecutive hospital stay records, and events for the same patients on the same days.
* Preparing a summary of events excluded from matching for any reason.

## Comparison of Data Distributions with Previous Years

Codebooks are created that allow comparisons of the most recent year data distribution with the distributions from the 2 prior years for all variables on all delivered files. No formal statistical tests are conducted; however, any apparent discrepancies between the most recent year distribution and previous years are explored. This check is particularly important to ensuring consistency in the variable definitions and resulting distributions between Medical Expenditure Panel Survey years.

# Summary of Key 2022 MPC Medical File Variables

Key variables from the Hospital, office-based doctor (OBD), and separately billing doctor (SBD) files are the payment and charge variables, and variables that give the reason for discrepancy between the payments and charges. As part of the quality control (QC) processes, RTI examines and compares the distributions of these variables with data from most recent 3-year period. This section summarizes the findings from this comparison.

RTI prepares summary tables that examine these variables across the most recent 3-year period, by type of event (OBD, total Hospital, Hospital emergency room, Hospital outpatient, Hospital inpatient, and institutional stays), mode of data collection (telephone or abstraction), and source of payment. Tabulations in this matching and editing plan are prepared separately for fee-for-service (FFS) and capitated events. Tables containing the summary statistics are produced. Hospital events provided by the VA are excluded from the exhibits in this section.

These distributions and summary statistics are visually compared to look for anomalies or outliers in the 2022 data. **Exhibits 5.1** and **5.2** display summary statistics for charges and payments from the Medical File data for FFS OBD events and for FFS Hospital events for the most recent 3-year period. As shown in **Exhibit 5.1**, in 2022, 44.2 percent of FFS OBD events had payments greater than $0 from private insurance, and 55.6 percent of events had payments equal to $0 from private insurance. Only 0.2 percent of events had missing or unknown values for this payment type and the average dollar amount paid by private insurance for each OBD event was $202.

**Exhibits 5.3** and **5.4** examine the reasons for discrepancy between total charge and total payment for the FFS OBD events and for the FFS Hospital events for the most recent 3-year period. For example, from **Exhibit 5.3**, of OBD FFS events where payments were less than charges and “Adjustments (DISADJ)” was coded as “yes” or “no” for the reason, 97.8 percent of the events have a response of “yes” that this is the reason.

Exhibit 5.1 Comparing Payments and Charge Variables Among Fee-for-Service OBD Events, 2020–2022 MPC⎯OBD (MV Only)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Payment/Charge | Percent > 0 | | | Percent = 0 | | | Percent =  Missing, Refusal, etc. | | | Mean Among Values > 0 | | |
| 2020 | 2021 | 2022 | 2020 | 2021 | 2022 | 2020 | 2021 | 2022 | 2020 | 2021 | 2022 |
| Payment from Patient | 33.3 | 33.4 | 37.6 | 66.4 | 66.3 | 62.2 | 0.3 | 0.3 | 0.2 | $87 | $84 | $92 |
| Payment from Medicare | 40.0 | 41.1 | 41.0 | 59.7 | 58.7 | 58.8 | 0.3 | 0.2 | 0.2 | $172 | $190 | $202 |
| Payment from Medicaid | 11.9 | 13.3 | 11.4 | 87.9 | 86.5 | 88.5 | 0.3 | 0.2 | 0.2 | $128 | $136 | $134 |
| Payment from Private Insurance | 48.0 | 41.8 | 44.2 | 51.8 | 57.9 | 55.6 | 0.3 | 0.3 | 0.2 | $213 | $201 | $202 |
| Payment from VA/CHAMPVA | 0.4 | 0.5 | 0.6 | 99.4 | 99.3 | 99.2 | 0.2 | 0.2 | 0.2 | $144 | $290 | $292 |
| Payment from TRICARE | 2.4 | 2.3 | 2.4 | 97.4 | 97.5 | 97.4 | 0.2 | 0.2 | 0.2 | $117 | $114 | $160 |
| Payment from Workers Comp | 0.5 | 0.3 | 0.4 | 99.3 | 99.5 | 99.4 | 0.2 | 0.2 | 0.2 | $279 | $402 | $322 |
| Total Payment | 95.4 | 94.8 | 95.5 | 4.2 | 4.9 | 4.2 | 0.3 | 0.3 | 0.3 | $233 | $235 | $248 |
| Total Charges | 99.7 | 99.8 | 99.9 | 0.0 | 0.1 | 0.0 | 0.3 | 0.1 | 0.1 | $517 | $555 | $588 |

CHAMPVA = Civilian Health and Medical Program of the Department of Veterans Affairs; MPC = Medical Provider Component; MV = medical visit; OBD = office-based doctor; SBD = separately billing doctor; VA = Veterans Affairs.

Note: The denominator of the percentages is all fee-for-service OBD events. The statistics are based on unique (nonreplicated) MPC events.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

Exhibit 5.2 Comparing Payments and Charge Variables Among Fee-for-Service Hospital Events, 2020–2022 MPC—Hospital All Events

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Payment/Charge | Percent > 0 | | | Percent = 0 | | | Percent =  Missing, Refusal, etc. | | | Mean Among Values > 0 | | |
| 2020 | 2021 | 2022 | 2020 | 2021 | 2022 | 2020 | 2021 | 2022 | 2020 | 2021 | 2022 |
| Payment from Patient | 21.1 | 21.7 | 23.6 | 78.7 | 78.1 | 75.8 | 0.2 | 0.2 | 0.7 | $219 | $200 | $219 |
| Payment from Medicare | 44.8 | 45.7 | 43.7 | 54.7 | 53.9 | 55.3 | 0.5 | 0.5 | 1.0 | $1,161 | $1,014 | $1,012 |
| Payment from Medicaid | 16.2 | 17.4 | 15.9 | 83.4 | 82.3 | 83.3 | 0.4 | 0.3 | 0.7 | $672 | $614 | $657 |
| Payment from Private Insurance | 35.0 | 31.1 | 35.5 | 64.8 | 68.7 | 63.6 | 0.2 | 0.3 | 0.9 | $1,316 | $1,176 | $1,295 |
| Payment from VA/CHAMPVA | 1.0 | 0.9 | 0.8 | 98.8 | 98.9 | 98.6 | 0.2 | 0.2 | 0.6 | $1,293 | $1,456 | $1,533 |
| Payment from TRICARE | 2.5 | 1.9 | 1.8 | 97.3 | 97.8 | 97.6 | 0.2 | 0.3 | 0.6 | $354 | $358 | $437 |
| Payment from Workers Comp | 0.4 | 0.4 | 0.5 | 99.4 | 99.5 | 98.9 | 0.2 | 0.2 | 0.6 | $1,914 | $1,094 | $961 |
| Total Payment | 89.1 | 90.2 | 89.5 | 10.2 | 9.1 | 9.3 | 0.7 | 0.7 | 1.2 | $1,362 | $1,178 | $1,272 |
| Total Charges | 96.9 | 96.9 | 97.1 | 2.9 | 2.8 | 2.4 | 0.2 | 0.3 | 0.5 | $5,036 | $4,311 | $4,341 |

CHAMPVA = Civilian Health and Medical Program of the Department of Veterans Affairs; MPC = Medical Provider Component; VA = Veterans Affairs.

Note: The denominator of the percentages is all fee-for-service office-based doctor events. These statistics are based on unique (nonreplicated) MPC events. Hospital events provided by the VA are not included.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

Exhibit 5.3 Comparing Reasons for Charge/Payment Discrepancy Among Fee-for-Service OBD Events, 2020–2022 MPC⎯OBD (MV Only)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Payment/Charge | Percent = Yes  (Among Events With  Yes or No) | | | Percent = Missing  (Among Events That Should Have a Response) | | |
| 2020 | 2021 | 2022 | 2020 | 2021 | 2022 |
| Payments < Charges, Adjustments (DISADJ) | 97.9 | 97.6 | 97.8 | 0.3 | 0.2 | 0.1 |
| Payments < Charges, Expecting Additional Payment (MOREPAY) | 4.8 | 4.5 | 5.1 | 0.3 | 0.2 | 0.1 |
| Payments < Charges, Charity (SLIDSCA2) | 1.5 | 1.4 | 1.4 | 0.3 | 0.2 | 0.1 |
| Payments < Charges, Bad Debt (BADDEB2) | 0.5 | 0.7 | 0.6 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Patient (EPAYPAT) | 3.1 | 2.7 | 3.1 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Medicare (EPAYCAR) | 0.6 | 0.8 | 0.6 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Medicaid (EPAYAID) | 0.4 | 0.6 | 0.5 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Private Insurance (EPAYPINS) | 0.8 | 0.8 | 1.0 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, VA/CHAMPVA (EPAYVA) | 0.0 | 0.1 | 0.1 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, TRICARE (EPAYCHAM) | 0.1 | 0.0 | 0.1 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Workers Comp (EPAYWORK) | 0.0 | 0.0 | 0.1 | 0.3 | 0.2 | 0.1 |
| Expect Additional Payment, Something Else (EPAYOTH) | 0.2 | 0.3 | 0.2 | 0.3 | 0.2 | 0.1 |

CHAMPVA = Civilian Health and Medical Program of the Department of Veterans Affairs; MPC = Medical Provider Component; MV = medical visit; OBD = office-based doctor; VA = Veterans Affairs.

Note: These statistics are based on unique (nonreplicated) MPC events.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

Exhibit 5.4 Comparing Reasons for Charge/Payment Discrepancy Among Fee-for-Service Hospital Events, 2020–2022 MPC—Hospital All Events

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Payment/Charge | Percent = Yes  (Among Events With  Yes or No) | | | Percent = Missing  (Among Events That Should Have a Response) | | |
| 2020 | 2021 | 2022 | 2020 | 2021 | 2022 |
| Payments < Charges, Adjustments (DISADJ) | 97.8 | 98.1 | 98.4 | 3.3 | 3.2 | 3.0 |
| Payments < Charges, Expecting Additional Payment (MOREPAY) | 7.9 | 7.2 | 7.9 | 3.3 | 3.2 | 3.0 |
| Payments < Charges, Charity (SLIDSCA2) | 1.2 | 1.1 | 1.2 | 3.3 | 3.2 | 3.0 |
| Payments < Charges, Bad Debt (BADDEB2) | 0.9 | 0.9 | 0.9 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Patient (EPAYPAT) | 5.5 | 4.7 | 5.6 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Medicare (EPAYCAR) | 0.9 | 1.0 | 0.8 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Medicaid (EPAYAID) | 0.7 | 0.9 | 0.6 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Private Insurance (EPAYPINS) | 0.9 | 0.9 | 1.0 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, VA/CHAMPVA (EPAYVA) | 0.1 | 0.0 | 0.0 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, TRICARE (EPAYCHAM) | 0.0 | 0.1 | 0.1 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Workers Comp (EPAYWORK) | 0.0 | 0.0 | 0.1 | 3.3 | 3.2 | 3.0 |
| Expect Additional Payment, Something Else (EPAYOTH) | 0.5 | 0.6 | 0.4 | 3.3 | 3.2 | 3.0 |

CHAMPVA = Civilian Health and Medical Program of the Department of Veterans Affairs; MPC = Medical Provider Component; VA = Veterans Affairs.

Note: These statistics are based on unique (nonreplicated) MPC events. Hospital events provided by the VA are not included.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

*[This page intentionally left blank.]*

# 2022 Medical File Pre-Matching Activities

Matching the Household Component (HC) and Medical Provider Component (MPC) data occurs after the data are edited. Matching requires properly formatted data files of medical events by person–provider pairs. One file comes from the contractor responsible for collecting the HC data. RTI constructs the other file, which contains MPC data. The files are

* Office-based doctor data (OBD) and Hospital data from the MEPS HC—the pre-matching HC data file constructed by the HC data collection contractor, which contains data on hospital events (inpatient Hospital stay, outpatient [OP], and emergency room [ER] events), office-based medical visits, and institutional care events, and
* Edited OBD and Hospital data from the MEPS MPC—the Medical events data file constructed by RTI by concatenating the files that contains the Hospital, OBD, and institution data

Similar files for Home Health (HH) data are also obtained and constructed. Medical data files and the HH files are matched separately. The HH data are separated from the Medical File data because the type of data known about the medical events differs (i.e., the variables that can be used during the matching process are different between the two pairs of files).

The goal in the matching task is to match records between the HC and MPC data (the two files listed above). Because not all medical events reported in the HC are eligible for inclusion in the MPC, ineligible events are excluded from the HC data files. The pre-matching activities for the MPC-HC are described in detail in the following sections.

## Obtaining the Household Survey Data

The HC datasets include various data from the HC survey data, including medical conditions; ER visits; inpatient hospital stays (HSs); OP visits; office-based medical visits; HH care events; data files containing provider, event, and patient identification numbers; and event and condition linkage variables. The goal is to begin with a HC data file that contains “clean,” edited data for only those person–provider pairs that responded to the MPC portion of the study.

## Obtaining the MPC Data from the Edited Files

RTI creates a SAS dataset from the MPC data to use in matching. Fields of critical importance in the matching, such as medical conditions, treatment procedures, and dates of events, are recoded to be complete and consistent with the corresponding fields in the HC datasets. Global fee and capitated event indicators are checked. For global fee or capitated events, only those expenses that occur in the most recent data collection year are included in the final dataset.

Dates of events are also examined. Hospital inpatient events and institution events that ended after the most recent data collection year (including events in which the patient was not yet discharged at the time of data collection) were excluded from the set used for matching prior to the 2020 MPC. Since the 2020 matching, MPC inpatient and institution events that ended by January 14 of the following year are eligible for matching and rolling with other inpatient, OP, or ER events. Other Hospital OP events, ER events, and medical visit (MV) events with an event date not in the most recent data collection year are excluded from matching. Charges and payment fields for MPC event are examined, and events that are missing both charge and payment data are excluded from matching.

## Medical File Rollup Groups

A **rollup group** is a set of events that *could* be considered a single event for matching. Separate events could be reported as one event in the HC file, and the rolling of medical events together should enable more to be matched. For example, on the HC file, an ER visit and subsequent inpatient stay might be reported as a single event. Rolling up these two events together on the Medical File increases the accuracy of a match. This section describes the rollup groups that were created for the MPC and how they are used in the matching task. Calculations of the charges or payments for the rollup groups is not described in this report; the HC contractor that is responsible for the imputation adds together charges or payments for the rolled events according to their specifications.

### Definition of Rollup Groups

As in prior cycles, 10 different rollup groups were created independently from one another; therefore, an event can belong to more than one rollup group. Each event contains nine indicator variables named ROLLUP\_FLAG1, ROLLUP\_FLAG2, …, ROLLUP\_FLAG10, which take the value 1 if the event belongs to the ith rollup group, and 0 otherwise. Each event also contains ID variables ROLLUP\_MPSID1, ROLLUP\_MPSID2, …, ROLLUP\_MPSID10, which identify the MPSID of the master record in each rollup. The master record in a rollup group *i* can be identified on delivered files by checking the following condition:

ROLLUP\_FLAGi = 1 and MPSID=ROLLUP\_MPSIDi, for i=1, 2, …, 10.

**Exhibit 6.1** summarizes the 10 rollup groups. The **master rollup event** is the primary event in the rollup group. All other events that are not the master event are called **child events**.

ROLLUP\_FLAG10 and ROLLUP\_MPSID10 are created after matching. ROLLUP\_FLAG10 indicates that an unmatched MPC inpatient stay has been rolled with a matched MPC inpatient stay because the beginning or ending dates of the unmatched MPC event is within 60 days of the ending or beginning dates of the corresponding matched household event, or the unmatched MPC inpatient stay overlaps the household event.

Exhibit 6.1 Summary of the Rollup Groups Created for the 2022 MEPS Matching

| Rollup Group | Types of Events Included in Rollup Group | Description of Group | Event in the Rollup That is Considered the Master Rollup Event | Events That Are Eligible for Matching |
| --- | --- | --- | --- | --- |
| 1 | Hospital inpatient, or | Inpatient HS events with consecutive stay dates for the same person by provider1, or | Event with earliest date (if all events in the rollup are capitated or all events are FFS), or | All2 |
| Hospital inpatient and Institutional stays | * Institutional stay events and HS events with consecutive stay dates for the same person by provider * Stays with overlapping dates are manually reviewed by data collection managers and are not rolled | Capitated event with the earliest date (if both capitated and FFS events in the rollup) |
| 2 | Institutional stays | Institutional stay events with consecutive stays dates for the same person by provider | * Event with earliest date (if all events in the rollup are capitated or all events are FFS), or * Capitated event with the earliest date (if both capitated and FFS events in the rollup) | All2 |
| 3 | ER events and Hospital inpatient | * ER events that fall on the first day or preceding day of an inpatient HS are rolled with the HS * Rollup Group 3 excludes ER events where the person was admitted from the ER and the ER charge was included in the inpatient stay amount * A small number of ER events with more than 1-day gap before an inpatient stay were rolled with the inpatient stay in situations where the patient was classified as admitted to the hospital from observation | HS | All (both ER and HS) |

(continued)

Exhibit 6.1 Summary of the Rollup Groups Created for the 2022 MEPS Matching (continued)

| Rollup Group | Types of Events Included in Rollup Group | Description of Group | Event in the Rollup That Is Considered the Master Rollup Event | Events That Are Eligible for Matching |
| --- | --- | --- | --- | --- |
| 4 | MV and OP visits | Multiple MV and OP events on the same day (there must be at least one MV and at least one OP to be a part of Rollup Group 4)  Note: beginning with the 2020 MPC, Rollup Group 4 excludes VA events. | * One of the capitated events (if both capitated and FFS events in the rollup), otherwise * One of the events * Event with earliest date (if all events in the rollup are capitated or all events are FFS), or * Capitated event with the earliest date (if both capitated and FFS events in the rollup). | All |
| 5 | Hospital inpatient | Inpatient HS events with stay periods that are 1 day apart (i.e., the end date of one stay and the begin day of the next stay have a 1-day gap) |  | All |
| 6 | Hospital OP and Hospital inpatient | One or more hospital OP events for the same person fall in the period of an inpatient HS (excluding the first day of the stay but including the last day of the HS) (either same or different providers)  Rollup Group 6 allows one or more embedded events and excludes the first day of the stay. This change, made in 2014, allows more events to be rolled, and prevents the same OP/HS inpatient events from being a part of Rollup Group 7. | HS | Master event (HS) |
| 7 | Hospital OP and Hospital inpatient | Hospital OP events that occur on the first day of an inpatient stay or the day before an inpatient stay | HS | All events (HS and OP) |

(continued)

Exhibit 6.1 Summary of the Rollup Groups Created for the 2022 MEPS Matching (continued)

| Rollup Group | Types of Events Included in Rollup Group | Description of Group | Event in the Rollup That Is Considered the Master Rollup Event | Events That Are Eligible for Matching |
| --- | --- | --- | --- | --- |
| 8 | Hospital ER and Hospital OP | Hospital ER events that occur on the same day as one or more OP events | * ER event (if all events in the rollup are capitated or all events are FFS), or * Capitated ER event (if there are both capitated and FFS events in the rollup), or * One of the capitated events (if there are both capitated and FFS events in the rollup). | All events |
| 9 | ER events, or | ER events for the same person on the same day with the same provider, or | Capitated event | Master event |
| MV events, or | MV events for the same person on the same day with the same provider, or |
| OP events | * OP events for the same person on the same day with the same provider. * Events of different provider types are not rolled as a part of this rollup group. |

(continued)

Exhibit 6.1 Summary of the Rollup Groups Created for the 2022 MEPS Matching (continued)

| Rollup Group | Types of Events Included in Rollup Group | Description of Group | Event in the Rollup That Is Considered the Master Rollup Event | Events That Are Eligible for Matching |
| --- | --- | --- | --- | --- |
| 10 | HS events | * Unmatched MPC inpatient events are rolled with matched MPC inpatient events when the beginning or ending date of the unmatched inpatient event is within 60 days of the ending or beginning date of a matched HC inpatient stay or the unmatched MPC inpatient stay overlaps the dates of a matched HC inpatient stay. * Rollup Group 10 was new for the 2014 MPC. * MPC inpatient events that match an HC OP event were disallowed from rolling with unmatched MPC inpatient events beginning with the 2022 cycle. | Matched MPC inpatient stay | NA |

ER = emergency room; FFS = fee for service; HC = Household Component; HS = hospital stay; MPC = Medical Provider Component; MV = medical visit; OP = outpatient.

1 Starting in 2020, the collapsed contact group used for matching (GID2) was used to define the “provider” for the rollup types.

2 Starting in 2019, all events in Rollup Groups 1 and 2 are eligible for the MPC-HC matching; prior to 2019 only the master event was eligible for matching.

Note: The number of events in the rollup groups are unique (nonreplicated) MPC events (including events provided by the VA).

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

### Variables Created for Selected Rollup Groups

### After the rollup groups are constructed, selected variables are created that identify rollup group linkages for some of the groups. Compound Rollups

Events can belong to more than one rollup group. A compound rollup group is one that is formed by combining multiple rollup groups with events in common. An event can be a child in multiple rollup groups (i.e., it could have multiple master events), or an event can be a master in one rollup group and a child in another rollup group.

If the event is a child in multiple rollup groups the “original” master event, MPSIDOLD, is created and defined to identify the single master event. For child events, MPSIDOLD is set to the MPSID of the original master event. The original master event is the one identified in the ROLLUP\_MPSIDi variables (where i represents each of the rollup groups 1 to 10). For example, the original master event for a child of rollup type 6 is in the variable ROLLUP\_MPSID6. When an event is a child in multiple rollup groups, the master is chosen from exactly one of the groups to serve as the original master event. **Exhibit 6.2** shows the source of MPSIDOLD when an event is a child in multiple rollup groups. In general, the source of MPSIDOLD is the ID whose master was an HS event, ER event, OP event, or MV event, in that order. The variable MPSIDOLD identifies the “original” master event; the term “original” is used to distinguish it from the “ultimate” master event, which is described below. This hierarchy is not relevant to inpatient stays in rollup group 1 (consecutive inpatient or institution/inpatient stays) and rollup group 5 (inpatient stays with a 1-day gap) because by the definition of these rollup groups, an inpatient stay would not be a child in both. It is possible that an institution stay could be a child in both rollup group 1 (consecutive inpatient or institution/inpatient stays) and rollup group 2 (consecutive institution stays), but this is rare and has not yet occurred in the data; if this occurred, the source would be the group with the earliest event.

Exhibit 6.2 Rollup Combinations

| Rollup Combination | Source of MPSIDOLD |
| --- | --- |
| 4, 6 | ROLLUP\_MPSID6 |
| 4, 7 | ROLLUP\_MPSID7 |
| 4, 6, 7 | ROLLUP\_MPSID6 |
| 6, 7 | ROLLUP\_MPSID7 |
| 7, 8 | ROLLUP\_MPSID7 |
| 6, 7, 8 | ROLLUP\_MPSID6 |
| 4, 8 | ROLLUP\_MPSID8 |
| 4, 7, 8 | ROLLUP\_MPSID7 |
| 3, 8 | ROLLUP\_MPSID8 |
| 3, 9 | ROLLUP\_MPSID9 |
| 4, 9 | ROLLUP\_MPSID4 |
| 6, 8 | ROLLUP\_MPSID8 |
| 4, 8, 9 | ROLLUP\_MPSID4 |
| 8, 9 | ROLLUP\_MPSID8 |
| 7, 9 | ROLLUP\_MPSID7 |
| 7, 8, 9 | ROLLUP\_MPSID7 |
| 6, 7, 8, 9 | ROLLUP\_MPSID7 |
| 3, 8, 9 | ROLLUP\_MPSID3 |
| 4, 7, 8, 9 | ROLLUP\_MPSID7 |
| 1, 2 | ROLLUP\_MPSID1 |
| 6, 7, 9 | ROLLUP\_MPSID6 |
| 6, 9 | ROLLUP\_MPSID6 |

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

Note: Rollup group number descriptions are shown in **Exhibit 6.1**.

**Exhibit 6.3** provides an example (using dummy ID variables) of how this hierarchy works. Rollup group 4 consists of an MV with MPSID=4587 and an OP event with MPSID=4214; the value of ROLLUP\_MPSID4 is 4214. Rollup group 7 consists of the same OP event with MPSID=4214 and an inpatient event with MPSID=5208; the value of ROLLUP\_MPSID7 is 5208. Rollup groups 4 and 7 form a compound rollup group because the same OP event is in both. MPSIDOLD is assigned the value in ROLLUP\_MPSID7 (which contains the inpatient stay) and MPSIDRLL takes the same value.

Exhibit 6.3 Example of Rollup Group Where Event Is a Child in More Than One Rollup Group

|  |  |  |  |
| --- | --- | --- | --- |
| Rollup Combination | ROLLUP\_MPSID | MPSIDOLD | MPSIDRLL |
| Rollup 4 and 7 | ROLLUP\_MPSID7=5208  ROLLUP\_MPSID4=4214 | 5208 | 5208 |
| Rollup 7 | ROLLUP\_MPSID7=5208 | 5208 | 5208 |
| Rollup 4 | ROLLUP\_MPSID4=4214 | 5208 | 5208 |

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

Another variable, MPSIDRLL, is defined for child events and is set to the MPSID of the “ultimate” master event. The ultimate master event considers the possibility that a child’s master may itself be a child in another rollup group, the master in that group may itself be a child in another group, etc. **Exhibit 6.4** is an example where the ultimate master event is different from the original event.

Exhibit 6.4 Example of Compound Rollup Group with an Ultimate Master Event Different Than Original Event

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rollup Type | Child Event | Master Event | MPSIDOLD | MPSIDRLL |
| 3 | A | B | B | D |
| 1 | B | C | C | D |
| 5 | C | D | D | D |

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

A choice for MPSIDRLL for event A could be event B. However, B is a child in a rollup group where C is the master, and C is a child in a rollup group where D is the master. So, As MPSIDRLL is set to D. Since D is not a child in any rollup group, the ultimate master event for A is D. Similar logic is used to set B and Cs MPSIDRLL to D also. In a sequence of rollup groups such as this, the ultimate master event is set to be the group whose master is not a child in any other rollup group.

An event can be a child in multiple rollup groups (i.e., it could have multiple master events); in this case, the “original” master event MPSIDOLD is used to identify the single master event. An iterative process is used to identify MPSIDRLL. Counts of the number of child events (i.e., not master events) in each rollup group are computed. MPSIDRLL is updated until it points to an event without child events.

At the conclusion of this process, MPSIDOLD is updated so that it contains a value only when that value differs from MPSIDRLL.

## Dealing with Rollup Groups During Matching

**Exhibit 6.1** shows that for some rollup groups, all events in the group are eligible for matching. For other rollup groups, only the master event is eligible for matching. Only the master event is eligible for matching in groups 6 and 9. All events are eligible for matching in the other groups. Starting in the 2019 MPC-HC matching, all events in groups 1 and 2 are eligible for matching; prior to 2019, the event with the earliest date was eligible for matching.

## Post-Matching Processing of Rollup Groups

After the matching of the MPC-HC Medical File data containing the individual events and rollup groups, MPSIDRLL is updated to reflect the results of the matching. A linkage file is prepared for delivery to the HC contractor. It consists of the union of the following records from Pass 1:

* Matched events, and
* Child events from rollup groups where the master event is a matched event. For Pass 1, all unmatched child events with a matched master event are included. For Pass 1.5, unmatched but match-eligible child events are not included.

The following outcomes of matching regarding the rollup groups can occur:

1. Simple events (i.e., an event not involved in a rollup group) that match.
2. Events in rollup groups where the master event matches but none of the child events match.
3. Excluding (2), all matched events in simple rollup groups (a simple rollup group is a rollup group not formed by combining multiple rollup groups).
4. Excluding (3), all events in simple rollup groups where a child matches and the master event did not match, but the definition of the rollup group allows the child to be reassigned as the master event (these are primarily events in rollup groups 4 and 5).
5. Excluding (2), all matched events in compound rollup groups (a compound rollup group is a rollup group formed by combining multiple rollup groups).
6. No events match.

For simple events [outcome (1)], MPSIDRLL is retained as the MPSID. All the simple events that match are placed on the linkage file.

For events in rollup groups where the master event matches but none of the child events match [outcome (2)], the MPSIDRLL is retained. The events in this rollup group that do not match are flagged. All events in rollup groups that fall into this category are placed on the linkage file.

In simple rollup groups where at least one event matches [outcome (3)], the matched events are added to the linkage file as simple events, and the MPSIDRLL variables are adjusted for events that do not match.

The linkage file also includes unmatched events that were a part of the simple rollup groups. This is accomplished by reassigning the MPSIDRLL for these unmatched events, when some but not all events in a simple rollup group are matched ([outcome (4)], implemented in 2015)

In compound rollup groups where at least one event matches [outcome (5)], it is more difficult to resolve the MPSIDRLL variable. In this case, each compound rollup group is processed iteratively to assign the MPSIDRLL variable, using the rollup types, which events matched, and the definitions for which types of events can be the master event. Unmatched events are reassigned a MPSIDRLL of a matched event in the same rollup group. This resolution is done with a computer algorithm, but the events are also output to a file for manual review and checking of the MPSIDRLL variable. The MPSIDRLL value for events that match is the MPSID of the event.

If no events match [outcome (6 above)], then the records are not added to the linkage file.

Since 2016, the matching process includes replicated MPC events from Pass 1.5 on the linkage file using the same rules as for Pass 1, with the exception that unmatched child events that are match eligible are not included on the linkage file even if their master event was matched at Pass 1.5.

Another adjustment after matching is that unmatched MPC inpatient events are rolled with matched MPC events for those unmatched MPC events within 60 days of the matching HC event. MPC events that overlap an HC event are also rolled with the corresponding matched MPC event. These events are denoted as “Rollup group type 10” events on the data files, which indicate that an unmatched inpatient stay was rolled with a matched stay. Unmatched MPC inpatient events were disallowed from rolling with MPC inpatient events that match HC Outpatient events in the 2022 MPC; this prevents multiple nights of inpatient stays from matching with an OP event and is consistent with the allowed and disallowed matches by event type. Rollup\_MPSID10 gives the MPSID of the inpatient stay with which the unmatched stay is rolled.

A set of variables added to the files beginning with the 2014 MPC denote when ER and HS events are linked. This resulted in the following changes:

* Rollup group 3 was modified to exclude ER/HS linked events.
* An ER/HS flag (variable name MERHSFLG) was created for all MPC events. MERHSFLG denotes the MPC ER event and associated hospital event where ADFROMER=1 and ERCHRGINC=1. When there is more than one ER event, the one on the same day is used. Variables ERHSFLAG, SQUARE, ERLINKID, HSLINKID were created for the HC events. For HC inpatient events, ERHSFLAG indicates whether the HC inpatient event is matched to a linked MPC inpatient event, and for HC events other than inpatient events (i.e., ER, OP, or MV) ERHSFLAG indicates that the HC event occurred between 2 days before or 1 day after the HC inpatient event, the event is matched to a linked MPC ER event, and the HC events within the HC linking pair are matched to MPC events within the same MPC link. The variable SQUARE indicates that the HC inpatient event is linked to an HC ER/OP/MV event that is matched to a linked MPC ER event. For HC inpatient events that are linked and have SQUARE=1, variables ERLINKID and HSLINKID give the HC EVPVID of the linked events.
* An additional file was created and delivered for use in the SBD processing. It contains the MPSID of the ER event and the corresponding MPSIDRLL. This file is to be used in later processing to assist in adding SBD amounts from unmatched MPC ERs to their linked inpatient events.

# 2022 MPC Statistical Matching (Record Linkage) Process

The ultimate product of this task is the linking of medical events between the Household Component (HC) and Medical Provider Component (MPC). Since unique identifiers do not exist for medical events that would allow us to simply merge “by” some set of variables, and because data items on both the HC and MPC are subject to recall and other forms of non-sampling error (e.g., the reported date of an event may not be accurate), the HC and MPC data are linked using a statistical, probabilistic matching process. The matching process is also referred to as record linkage or record matching in the literature. Various algorithms have been developed and successfully used to not only identify matched records between two or more files, but also to identify duplicates within a file.

The statistical matching process for the Medical Expenditure Panel Survey (MEPS) data entails using several fields (or variables) to construct a **match weight** for all possible matched pairs between the HC and MPC files. In the statistical matching literature, other terms such as “composite match weight,” “composite weight,” and “total match weight” are used interchangeably with “match weight” (e.g., Jaro, 1989). The variables used in the matching must be available on both the HC and MPC data files, although they can contain missing values for some events (generally missing values will result in a zero-field comparison score [or contribution for the variable] to the match weight.) The match weight is a function of the similarity between non-missing values of these variables, as well as a function of some agreement probabilities estimated at the variable level. The match weight represents the estimated likelihood that a match is correct where pairs with a larger match weight have a higher probability of being an accurate match.

This section summarizes the methodology of the record linkage matching process used in MEPS.

## The Matching Algorithm

The Fellegi-Sunter algorithm links records based on match weights that are computed from the specific fields (or variables) being compared. The matching algorithm accommodates discrepancies in responses in the matching fields being compared on two files and provides probabilistic linkages even when the link is not perfect. This method determines the strength, or the quality, of the specific matches made. The process consists of assigning a comparison score to each field for a pair based on the likelihood that the match is for the same event. Thus, matching events would have large agreement weights, and nonmatches would have much smaller agreement weights.

The method assumes matching fields are independent. The match weight is computed for each pair of records from a likelihood ratio based on the agreement of the matching fields. The SAS macros match on arrays of numeric variables and have not been adapted for matching of text or character strings.

Suppose the index *i* denotes a field or variable used in the comparison process. In addition, suppose there is interest in determining a match weight for a pair of records *j* (from the HC file) and *k* (from the MPC file). Also, suppose the following:

*mi* = Pr {field *i* agrees in the set of M true matches}

*ui* = Pr {field *i* agrees in the set of U true nonmatches}

Thus, *mi* is the probability that field *i* agrees among the true matched records and *ui* is the probability that field *i* agrees by chance among true nonmatched records. These are referred to as the “*m*” and “*u*” probabilities.

The **agreement field weight** for variable *i* is defined as the base-2 logarithm of the ratio of these two probabilities:

 (7.1)

In addition, the **disagreement field weight** for variable *i* is defined as:

 (7.2)

The **field comparison score** refers to either of these values:



For a pair of records *j* (from the HC file) and *k* (from the MPC file), the notation used is

 (7.3)

The term “field comparison score” has several other synonyms in the literature, including field agreement/disagreement weight, component weight, contribution of each field, weight contribution of each field, and individual weights for agreement or disagreement on each field (e.g., Jaro 1989).

In addition to sharing the fundamental matching algorithm:

* Values produced from Equations (7.1) and (7.2) can be overridden and the agreement and disagreement field weight set to some fixed number. For example, the field comparison score for the event beginning date field is set to a value based on a nonlinear function of the number of days between event start dates from the HC and MPC events under consideration.
* The agreement weight can be prorated by the extent of agreement if the field does not match exactly. This option is used with the number of stay days for an inpatient event.
* The **Array** feature is available. Arrays are sets of variables that are considered a single entity when computing a field comparison score. This feature is used with a set of indicators that flag whether an event includes surgery, radiology, and/or laboratory. It is also used with a set of condition code indicators[[2]](#footnote-2) derived from Internal Classification of Diseases (ICD) variables.

A **match weight** for a pair of records *(j, k)* is formed by summing the field comparison scores for all selected fields. That is, the match weight for a pair of records *(j, k)* is defined as

 (7.4)

Since  tends to be a larger (positive) number if field *i* agrees and tends to be a smaller (negative) number if field *i* disagrees for a pair of records *(j,k)*, a larger positive match weight indicates a higher likelihood of being a true matched pair. The match weight is therefore a measure of the likelihood that a pair of records from the HC and MPC files matches. Those event pairs with match weights greater than a preset threshold can be considered true matches. It also follows that those event pairs with match weights less than this threshold are declared nonmatches.

A threshold is a means to control linkage errors, specifically, false positive and false negative error rates. This control is specified by a macro variable in the matching software. Since 2012, the minimum threshold was set to -10,000 for both passes of matching (i.e., effectively no minimum, except to prevent certain types of events from matching).

RTI’s software also includes an algorithm that selects the best set of matches after the match weights for all possible pairs have been computed. This “best” set of matches is defined as the set of HC and MPC pairs that yields the maximum total match weight within a block group, defined as the sum of the match weights of all selected pairs. The problem of identifying the pairs that give this maximum, subject to the various constraints built into this problem (e.g., a record can only be included in one pair), is often referred to in the literature as the Linear Assignment Problem or the Linear Sum Assignment Problem. This is a special case of the well-known Transportation Problem in operations research.

## Matching in Two Passes

In the MEPS, is run twice to establish the best link between the HC and MPC events. In Pass 1 the patient and medical provider contact group identifiers are used to create block groups. Beginning with the 2016 matching, when there are unmatched HC events in a patient and medical provider block group remaining from Pass 1, the number of unmatched HC events were counted and the match-eligible MPC events within the same block group were replicated once for each unmatched HC event (up to a maximum of 5 times) and allowed to match with the remaining HC events. This matching pass that uses replicated MPC events is called “Pass 1.5.” MPC events and HC events that were part of a global fee, as well as MPC events that were linked to a global fee event through one of the rollup groups were not eligible for matching in Pass 1.5. The *mi*’s and *ui*’s are estimated separately for each pass.

# Variables Used in the 2022 Medical File Matching

This section discusses the variables (or fields) used to construct the composite match weight for the 2022 Medical Provider Component-Household Component (MPC-HC) record linkage activity. The match weight is defined in **Equation (8.1)** and is defined as the sum of field weights ’s for field *i*, HC event *j* and MPC event *k*.

Match Weight =  (8.1)

## Blocking Variables

**Exhibit 8.1** presents both the number and percentage of events eligible for matching in the most recent 3-year period. For the 2022 cycle, 57,837 HC events and 68,836 MPC events were eligible for matching. If match weights were to be constructed for all possible pairs, then the weights would need to be computed almost 4 billion MPC and HC pairs, specifically:

57,837 \* 68,836 =

Exhibit 8.1 Events Eligible for Matching, 2020–2022

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2020 MPC | | 2021 MPC | | 2022 MPC | |
| Total | Percent | Total | Percent | Total | Percent |
| HC Events on File | 174,464 |  | 207,966 |  | 151,440 |  |
| MPC Events on File | 87,226 |  | 91,660 |  | 81,291 |  |
| Match-Eligible HC Events | 58,868 | 33.7 | 63,302 | 30.4 | 57,837 | 38.2 |
| Match-Eligible MPC Events | 70,467 | 80.8 | 73,554 | 80.2 | 68,836 | 84.7 |

HC = Household Component; MPC = Medical Provider Component.

Note: A “match-eligible” event is an event that is eligible for the matching process. Events may be ineligible for a variety of reasons, such as the provider was in the HC sample but did not respond to the MPC, the event was declared a duplicate of another event in the files, and the date of the event indicated it was out-of-scope for the analysis. The counts are unique (nonreplicated) MPC events.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

This is computationally impractical. More importantly, considering all possible matches limits the effectiveness of the iterative algorithm used to derive the final match weights. To address both these issues, the MPC matching is done in two passes. The distinguishing feature of the two passes is the **blocking variables** used in the matching process. Blocking variables in the Medical Expenditure Panel Survey (MEPS) matching algorithm are variables used to partition the sample into non-overlapping groups, similarly to the way in which sampling strata are formed when samples are selected. At each iteration of the matching algorithm, the “best” match of MPC-HC records are identified within each block group. Agreement and disagreement field weights are computed for each field, and the field weights are used to compute a “better” match weight at the next iteration of the matching algorithm. The iterative process continues until differences between the agreement and disagreement field weights between iterations are less than a set threshold (0.001 is used for the MPC-HC matching). In the MPC, the following variables are used to define the blocking groups for the two passes.

**Pass 1 and Pass 1.5: Blocking Factor—Patient and Collapsed Contact Group**

In the two passes of the matching, the blocking groups are defined by patient and collapsed contact group. In the MPC, a **contact group** is a group of providers with the same point of contact (POC). The data collection staff (DCS) contacts this POC and asks the POC to provide data for all medical events for specific patients associated with any providers in the contact group. For the purposes of forming blocking groups, these contacts groups are further combined into **collapsed contact groups**. Contact groups are pooled together to collapsed contact groups if the provider names or other contact information are similar between any two providers that belong to different contact groups. Contact groups are collapsed in a manner that always preserves the initial contact group assignment; contact groups are not split to create collapsed contact groups for the blocking. Both HC and MPC provider information are used to create these collapsed contact groups. The goals of creating collapsed contact groups are to

* Combine providers with the same MPC POC to address any misclassification errors that occur if the MPC POC reports an event for a different provider in the same contact group.
* Combine providers with the same or similar name or other identifying information to address any misclassification errors that occur if the HC respondent reports an event for an incorrect provider.

## Construction of Collapsed Contact Groups for Matching in 2022

Collapsed contact groups are created to form the provider block groups that are used in matching. Specifically, provider name and address information from both the MPC and HC are used to create consistent, collapsed contact groups between the MPC and HC samples. The collapsed contact group used for blocking in the matching at Pass 1 and Pass 1.5 is the variable **GID2** and is formed using **GID**, the ID of the data collection contact group, as the starting point.

The HC events included in matching are those in collapsed contact groups that have one or more responding MPC providers in the group. This allows us to address differences in the assignment of provider ID from the HC survey. For example, if HC provider ID #1 and #2 refer to the same provider (i.e., they are in the same collapsed contact group), provider ID #1 has responding MPC records and provider #2 contains the HC provider events, then HC events associated with provider #2 were included in the matching process (assuming that provider #1 and provider #2 could be determined to actually be the same).

The MPC consists of a subsample of HC patient–provider pairs, and not all the providers identified in the HC are in the MPC sample. HC providers that were also in the MPC sample could be merged into the MPC sample directly (using the HC provider ID), and the GID could be assigned to place them into a contact group with MPC providers. The same HC provider can have different HC provider IDs, and HC events that match to collected MPC events sometimes exist under different HC provider IDs and with different spellings of the provider’s name or different telephone numbers. The HC providers whose provider IDs are not in the MPC sample are grouped, when possible, with the MPC sample providers using name, address, and telephone numbers.

Prior to the collapsing of contact groups, the provider name and street fields are standardized to account for variations in the use or format of abbreviations. For example, all occurrences of STREET are change to ST, and all occurrences of CENTER are changed to CTR.

Collapsed contact groups for blocking the combined file of HC and MPC providers for the 2022 matching were formed using the following sequence of steps for automatically grouping providers:

* The contact group (GID) from data collection was the starting point for forming the blocking variable. In the 2022 cycle, the GIDs for pairs that were disavowals were ignored and instead the provider was combined into GID2 using the grouping algorithm described later in this section.[[3]](#footnote-3)
* Providers for events that were programmatically resolved as duplicate events were placed in the same collapsed contact group.
* Providers were combined so that those with the same PROVID, PDDIRID, or GID have the same GID2.
* Patient–provider pairs are sometimes merged during data collection, but the contact groups are not merged (e.g., during reviews of duplicate events). For use in matching, RTI collapsed the contact groups associated with the merged pairs.
* The HC providers (PROVID) were combined with the MPC providers.
* HC providers that were identified as duplicates during the sample processing and prior to sample fielding were grouped into the same collapsed contact group. These were reviewed in the 2022 cycle and any that appeared to be different providers were not immediately combined into the same GID2. This should assist in preventing distinct providers with similar names (such as where the provider names contain the same geographic area) from being incorrectly placed into the same block group.
* Large providers were placed into their own GID2, using the data collection Band ID and text searches of the provider name and state to identify the providers.
* All MPC and HC providers were examined for persons with VA data to see if any additional HC providers should be placed in the VA GID2.
* Further collapsing of contact groups (GID2s) was accomplished using fuzzy matching. The GID2s were further combined into collapsed contact groups based on similar names, addresses, and telephone numbers. This collapsing helps allow the data from HC events whose pair was not in the MPC sample (and would therefore not have a GID) to be used in the matching. Starting in 2019, the steps used for matchings were reordered compared to prior years to improve the efficiency and to eliminate redundancies. For example, analyses from the 2016 – 2018 matching indicated that combining GID2s by telephone number accounts for most GID2s that are collapsed.

##### Review of Matching:

RTI conducted three types of manual review to look for additional matches and for over-grouping that may have resulted in erroneous matches. Those three types of manual review were:

(1) Inpatient, outpatient, OBD, and ER events for persons with unmatched MPC or HC hospital inpatient events after the first pass were examined to see if the unmatched events might have a matching event in a different GID2.

(2) Events for persons with unmatched MPC or HC events on the same day were also examined.

(3) Events for PERSID by GID2 sets that contained more than one GID were reviewed to identify sets where the providers were not in the same health care system as well as sets with relatively large numbers of HC events with no MPC events on the same date.

## Matching Fields (Variables)

The matching process used in the MPC creates match weights between all possible pairs within block groups. The match weight reflects the agreement and disagreement between a set of fields (variables) available on both the HC and MPC.

RTI and Agency for Healthcare Research and Quality (AHRQ) have conducted several reviews and tests of the coding and treatment of missing variables, the actual variables used in the matching, and the comparison score functions. In 2020, the MPC and HC collected information that indicates whether the OBD or Outpatient event is a telehealth visit. This “telehealth” indicator was included as a matching variable.

**Exhibit 8.4** presents all variables used to construct the match weight for the MPC. Some events reported by the HC respondent as inpatient stays appeared to be collected in the MPC as multi-night observation outpatient stays. However, the events would not match because of restrictions placed on the types of events that could match. Based on specifications from AHRQ for the 2017 MPC, a variable was created to flag MPC OP events that are possibly multiple-night Observational Stays, and these events were then allowed to match with HC-reported inpatient stays. This same definition has been used since 2017 for MPC-HC matching.

Exhibit 8.4 Matching Variables in 2022

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Use in the Construction of a Match Weight |
| BEGDATE\_SAS | Event begin date | Event dates (and event type) are the best indicators of a match between medical events in the HC and MPC. Because of the importance of event date, the contribution of the date to the match weight is set to be relatively large, and the contribution to the match weight increases in a nonlinear fashion as the difference approaches zero. The equation used to determine the field comparison score for event date was provided by AHRQ analysts. |
| EVENT | Event type. This variable is defined as:  1=ER  2=H1 (hospital stay, 1-–2 days)  3=H2 (hospital stay, 3 days)  4=H3 (hospital stay, 4+ days)  5=MV (medical visit)  6=OP (outpatient, no surgery, observation not indicated)  7=OS (outpatient, with surgery, observation not indicated)  9=MPC event, OP, multi-night observation indicated | EVENT is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of EVENT.  OP observation events refer to MPC Outpatient events where the total charge and CPT codes indicate that the event may have been a multi-night observation stay. |
| GLOFEE | Global fee indicator | GLOFEE is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of GLOFEE.  Starting with the 2018 MC-HC matching, global fee events, and events connected to global fee events through rollup groups, were not eligible for the replicated (Pass 1.5) matching. |
| HSPLUS1 | Length of hospital stay. Specifically, end date − begin date + 1. This is set to missing for non-inpatient events. | HSPLUS1 was considered a range variable when computing the match weight. |
| DAY | Day of month of event begin date. Takes values 1-31. This is set to missing when the event begin date is missing. | DAY is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of DAY.  DAY was not used at Pass 1.5 matching. |
| MONTH | Month of event begin date. Takes values 1-12. This is set to missing when the event begin date is missing. | MONTH is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of MONTH.  MONTH was not used at Pass 1.5 matching. |
| DAYWEEK | Day of week of event begin date. Takes values 1-7, with 1=Sunday, 2=Monday, etc. This is set to missing when the event begin date is missing. | DAYWEEK is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of DAYWEEK.  DAYWEEK was not used at Pass 1.5 matching. |

(continued)

Exhibit 8.4 Matching Variables in 2022 (continued)

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Use in the Construction of a Match Weight |
| ROUND | HC data collection round number of event begin date. This is set to 0 when the round for the MPC event could not be inferred from the dates of the HC rounds for the person. | ROUND is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of ROUND.  ROUND was not used for the Pass 1.5 matching. ROUND has values 1 – 9 in the 2021 MPC. |
| FLAG1, FLAG2, FLAG3 | Yes/no indicators for surgery, radiology, and laboratory (based on CPT/BETOS codes) | The three FLAG# variables were considered an array set when computing a match weight. |
| C700-C781, C797, C798, C799 | Yes/no condition indicators (recode of ICDCND) | The 85 C# variables were considered an array set when computing a match weight. |
| INPATIENT | Yes/no indicator if event type was an inpatient stay | INPATIENT is a categorical variable in the matching. Two records agree if they have the same responses and disagree if responses differ. Equations (7.1, 7.2, and 7.3) are used to determine the field comparison score of INPATIENT. |
| DONT1 | This variable was set equal to:  1=For hospital stays of 0,1 nights  2=For medical events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1. See Comment 4. |
| DONT2 | This variable was set equal to:  1=For hospital stays of 2 nights  2=For HC outpatient events (no surgery); MPC outpatient events without indication of observation (no surgery), medical visit or ER events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1. See Comment 4. |
| DONT3 | This variable was set equal to:  1=For hospital stays of 3+ nights  2=For HC outpatient events, MPC outpatient events without indication of observation, medical visit, or ER events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1. See Comment 4. |
| DONT4 | This variable was set equal to:  1=For MPC outpatient stays classified as observation stays  3=all other event types | This variable allowed the MPC outpatient stays to match with any of the HC event types at Pass 1, including inpatient events. |
| DONT5 | This variable was set equal to:  1=For HC MV events  2=For MPC ER events that are linked to any inpatient event  3=Otherwise | This variable prevents HC MV events from matching with MPC ER events that are linked to an inpatient stay at Pass 1. See Comment 4. |
| DONT6 | This variable was set equal to:  1=For HC OP events without surgery  2=For MPC ER events that are linked to inpatient events with 2 or more nights  3=Otherwise | This variable prevents HC OP events without surgery from matching with MPC ER events that are linked to inpatient stays with 2 or more nights at Pass 1. See Comment 4. |
| DONT7 | This variable was set equal to:  1=For HC OP events with surgery  2=For MPC ER events that are linked to inpatient events with 3 or more nights  3=Otherwise | This variable prevents HC OP events with surgery from matching with MPC ER events that are linked to inpatient stays of 3 or more nights at Pass 1. See Comment 4. |

Exhibit 8.4 Matching Variables in 2022 (continued)

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Use in the Construction of a Match Weight |
| DONT1a | This variable was set equal to:  1=For hospital stays of 0,1 nights  2=For medical visit, ER, or outpatient events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT2a | This variable was set equal to:  1=For hospital stays of 2 nights  2=For medical visit, ER, or outpatient events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT3a | This variable was set equal to:  1=For hospital stays of 3+ nights  2=For medical visit, ER, or outpatient events  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |

|  |  |  |
| --- | --- | --- |
| DONT4a | This variable was set equal to:  1=For outpatient events with surgery  2=For inpatient events, medical visit, ER (no surgery), or outpatient events (no surgery)  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT5a | This variable was set equal to:  1=For outpatient events (no surgery)  2=For inpatient events, medical visit, ER (with surgery), or outpatient events (with surgery)  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT6a | This variable was set equal to:  1=For ER events (with surgery)  2=For inpatient events, medical visit, ER (no surgery), or outpatient events (no surgery)  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT7a | This variable was set equal to:  1=For ER events (no surgery)  2=For inpatient events; Outpatient events with surgery; and ER events with Surgery  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT8a | This variable was set equal to:  1=For MV events  2=For inpatient events; Outpatient events with surgery; and ER events with Surgery  3=Otherwise | This variable was used to prevent certain types of events from matching in Pass 1.5. See Comment 4. |
| DONT9a | This variable was set equal to:  1=For HC MV events  2=For MPC ER events that are linked to any inpatient event  3=Otherwise | This variable prevents HC MV events from matching with MPC ER events that are linked to an inpatient stay in Pass 1.5. See Comment 4. |
| DONT10a | This variable was set equal to:  1=For HC OP events without surgery  2=For MPC ER events that are linked to inpatient events with 2 or more nights  3=Otherwise | This variable prevents HC OP events without surgery from matching with MPC ER events that are linked to inpatient stays with 2 or more nights in Pass 1.5. See Comment 4. |

Exhibit 8.4 Matching Variables in 2022 (continued)

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Use in the Construction of a Match Weight |

|  |  |  |
| --- | --- | --- |
| DONT11a | This variable was set equal to:  1=For HC OP events with surgery  2=For MPC ER events that are linked to inpatient events with 3 or more nights  3=Otherwise | This variable prevents HC OP events with surgery from matching with MPC ER events that are linked to inpatient stays of 3 or more nights in Pass 1.5. See Comment 4. |
| FORCE1 | This variable is set equal to 1, 2, 3… and is used to force a match between an MPC-HC pair within a block group. | This variable can be used to force certain events to match. See Comment 5.  FORCE1 was not used for the 2022 matching. |

AHRQ = Agency for Healthcare Research and Quality; BETOS = Berenson-Eggers Type of Service; CPT = Current Procedural Terminology; ER = emergency room; HC = Household Component; MPC = Medical Provider Component; MV = medical visit; OBD = office-based doctor; OP = outpatient; SBD = separately billing doctor.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

The variable OP\_OBS is set to 1 for MPC Hospital outpatient events for either of the following scenarios:

* Any of the following text strings appear anywhere in any of the MCPT\*, ORIG\_MCPT\*, and ORIG\_MCPTDS\* fields (regardless of field length/position in field): “99217,” “99218,” “99219,” “99224,” “99226,” “99234,” “99235,” “99236,” “obs,” “OBS,” “Obs,” “Obv,” “obv,” “OBV,” “g0378,” “G0378” AND the outpatient total charge (OUT\_TLCHRG) >8000 (~94 percentile) AND OUT\_TLCHRG is not missing
* OUT\_TLCHRG>16000 (~97-98 percentile) AND OUT\_TLCHRG is not missing

The variable EVENT is set to 9 when OP\_OBS=1. For the MPC, the variable EVENT takes values 1-9:

1 = ER

2 = Hospital Stay, 1 or 2 days

3 = Hospital Stay, 3 days (2 nights)

4 = Hospital Stay, 4 + days (3 + nights)

5 = Medical Office Visit

6 = Outpatient, no surgery, outpatient observation not indicated

7 = Outpatient, surgery, outpatient observation not indicated

8 = Home Health

9 = Outpatient, outpatient observation indicated.

For the HC, the variable EVENT is not changed from prior years and takes values 1-8:

1 = ER

2 = Hospital Stay, 1 or 2 days

3 = Hospital Stay, 3 days (2 nights)

4 = Hospital Stay, 4 + days (3 + nights)

5 = Medical Office Visit

6 = Outpatient, no surgery

7 = Outpatient, surgery

8 = Home Health

At Pass 1, MPC events with EVENT=9 (or OP\_OBS=1) were then allowed to match with HC inpatient hospital stays with EVENT=3 or 4 (as well as EVENT=1, 2, 5, 6, or 7 as previously). Pass 1.5 was more restrictive on the types of events that can match.

**Comment 1: Contribution of Begin Date to the Match Weight**

For matching, event date is treated as a range variable and does not simply agree or disagree. Event date is the most important variable used to identify a match between medical events in the HC and MPC. Given its importance, the contribution of event date is set to be relatively large and dependent on the absolute difference in the dates between the HC and MPC events under consideration. The difference between event dates is a continuous variable; as the difference in dates approaches zero, the contribution of the date to the match weight becomes larger in a nonlinear fashion. The function used since 2013 is:



where

= absolute difference in the event begin date between HC event j and MPC event *k*

= scale = steepness of decay (greater value = steeper decay and is hard-coded to a constant 4.5)

= minimum (the disagreement score for the event begin date or user-provided value), set to   
 -1 \* agreement score

= range (maximum – , where “maximum” is the agreement score or user-provided value

**Comment 2: Contribution of HSPLUS1 to the Match Weight**

The comparison score for length of stay that has been used since the 2013 matching is the following function:



where

= minimum = disagreement score for HSPLUS1

= range (agreement score - disagreement score for HSPLUS1)

= absolute difference between length of stay for HC event *j* and MPC event *k*

*HCLOSj* = Length of stay for HC event *j*

= width factor, flatness of decay (greater value = flatter decay) = 1.4

Starting in 2016, the inpatient stay matches at Pass 1.5 are further restricted so that the difference in the length of stay is less than 3 days.

**Comment 3: Array Variables**

Two sets of array variables were used in the MPC-HC matching. One set includes the surgery, radiology, and laboratory indicator variables {FLAG1, FLAG2 and FLAG3} respectively. A second set includes the condition indicator variables: {C700, C701, …, C799}.

When the software computes a match weight, array variables that belong to the same set are treated as a single entity, and the field comparison score (or contribution to the match weight) of the set is based on the agreement and disagreement of the variables within the set. Specifically, if an HC and MPC event have a response that agrees for one or more of the variables in an array set, then the contribution to the match weight from this set of variables is equal to the average agreement weights among variables in the set that agree. Otherwise, the match weight contribution is set equal to the average disagreement weight among variables in the set. For the surgery, radiology, and laboratory indicator variables individually, if both an HC event has “yes” and an MPC event has “yes,” or both the HC event and the MPC event say “no,” then the events are said to agree. If either the HC event or the MPC event has “yes,” but the other has “no,” then the events are said to disagree on the indicator variables. The field comparison score is calculated for each of the three variables separately, using equation 7.1 or 7.2; **Exhibit 8.5** shows how the field comparison score is assigned for values of the MPC and HC variables. Missing values in these three indicator variables are set to “no,” so there is no missing data. When disagreement occurs, the HC value is used in computing equation 7.2. When at least one of the three variables agree, the field comparison score for the set of variables is computed as the mean agreement score; otherwise, when none agree, the field comparison score is the mean disagreement score. **Exhibit 8.5** shows the comparison score matrix for determining agreement.

Exhibit 8.5 Comparison Score Assigned for Values of the MPC and HC Surgery, Radiology, or Laboratory Flags

|  |  |  |
| --- | --- | --- |
| HC Response to Surgery, Radiology, or Laboratory Indicator | MPC Response to Surgery, Radiology, or Laboratory Indicator | |
| Yes | No |
| Yes | Agreement score | Disagreement score |
| No | Disagreement score | Agreement score |

HC = Household Component; MPC = Medical Provider Component.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

For the array of condition indicators, the agreement or disagreement weight is computed separately for each of the 85 conditions. If the HC event and the MPC both are “yes” for a condition, then they “agree” for the condition, and the agreement weight is calculated using equation 7.1. If one is “yes” for a condition and the other is “no,” then they “disagree” for the condition and the disagreement weight is calculated using equation 7.2, using the HC value of the condition indicator for computing the value of equation 7.2. When the condition indicator for both the HC and MPC events is “no,” or is missing for one or both events, the events are treated as neither agreeing nor disagreeing, and the agreement and disagreement field weights are set to 0. This is shown in the **Exhibit 8.6**.

Exhibit 8.6 Comparison Score Assigned for Values of the MPC and HC Condition Indicators

|  |  |  |  |
| --- | --- | --- | --- |
| HC Response to Condition Indicator | MPC Response to Condition Indicator | | |
| Yes | No | Missing |
| Yes | Agreement score | Disagreement score | 0 |
| No | Disagreement score | 0 | 0 |
| Missing | 0 | 0 | 0 |

HC = Household Component; MPC = Medical Provider Component.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

Note: Field comparison score for the condition is set to 0 when the events neither agree nor disagree.

The field comparison score for the array of condition indicators is calculated from the individual agreement and disagreement weights as follows: if there is agreement for at least one of the conditions, then the contribution is the average of the agreement weights for those that agree; otherwise, it is the average of the disagreement weights for those that disagree.

**Comment 4: DONT1, DONT2, DONT3, DONT4, DONT5, DONT6, DONT7 Variables and DONT1a, DONT2a, DONT3a, DONT4a, DONT5a, DONT6a, DONT7a, DONT8a, DONT9a, DONT10a, and DONT11a Variables**

Historically, the MPC has not allowed matches between certain types of events between the HC and MPC. Those matches that are allowed and not allowed at Pass 1 are summarized in **Exhibit 8.7**. This exhibit indicates, for example, that an inpatient stay (HS) of 0 or 1 night on one of the two files cannot be paired up to a medical visit (MV) on the other file. As another example, an inpatient stay of 2 nights is not allowed to match with OP no surgery, ER, or medical visit event on the other file.

Pass 1.5 is more restrictive in that inpatient stays can only match other inpatient stays, matched inpatient stays must have a difference in length of stay of less than 3 days, and an event with surgery can only be paired with another event with surgery, as shown in **Exhibit 8.8**.

As noted in the previous section, beginning with the 2009 MPC, all events within a block group from one of the files are matched, with the exception that certain types of events are not allowed to match. To accomplish this goal of accepting as many allowed matches as possible within a block group in the matching process, the threshold, or a minimum allowable weight for any pair, is set to an extremely small value (specifically -100,000) and assigned the match weight to any unallowable pair to be this minimum threshold times 10,000 (any number less than the threshold of -100,000 could have been used). The variables DONT1 through DONT7 provide a method to algorithmically identify unallowable matches for Pass 1. For example, within a block group, if an HC event had a value of DONT1=1 and an MPC event had a value of DONT1=2, then the pair would receive a match weight of -100,000\*10,000. The variables DONT1a through DONT11a serve a similar purpose for Pass 1.5.

**Exhibit 8.4** shows how DONT1 through DONT7 are defined as well as how DONT1a through DONT11a are defined.

Exhibit 8.7 MPC-HC Matches Allowed and Not Allowed at Pass 1, 2022

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MPC Classification** | **HC Classification** | | | | | | | | |
| **HS, 0, 1 Night** | **HS, 2 Nights** | **HS, 3 Nights** | **OP, With Surgery** | **OP, No Surgery** | **ER, With Surgery** | **ER, No Surgery** | **MV** | **Total** |
| Hospital Stay, 0, 1 Night | + (106) | + (34) | + (22) | + (6) | + (7) | + (0) | + (1) | X | 176 |
| Hospital Stay, 2 Nights | + (18) | + (146) | + (62) | + (0) | X | X | X | X | 226 |
| Hospital Stay, 3 Nights | + (7) | + (19) | + (514) | X | X | X | X | X | 540 |
| Outpatient, With Surgery, multi-night observation not indicated | + (5) | + (3) | X | + (417) | + (588) | + (4) | + (22) | + (864) | 1,903 |
| Outpatient, No Surgery, multi-night observation not indicated | + (42) | X | X | + (299) | + (8,926) | + (9) | + (293) | + (9,039) | 18,608 |
| Outpatient, With Surgery, multi-night observation indicated | + (52) | + (2) | + (6) | + (213) | + (74) | + (6) | + (1) | + (70) | 424 |
| Outpatient, No Surgery, multi-night observation indicated | + (17) | + (3) | + (1) | + (9) | + (98) | + (0) | + (1) | + (48) | 177 |
| Emergency Room With Surgery linked to <2 night Inpatient | + (0) | X | X | + (0) | + (0) | + (0) | + (0) | X | 0 |
| Emergency Room No Surgery linked to <2 night Inpatient | + (0) | X | X | + (2) | + (8) | + (7) | + (68) | X | 85 |
| Emergency Room With Surgery linked to 2 night Inpatient | + (0) | X | X | + (0) | X | + (0) | + (0) | X | 0 |
| Emergency Room No Surgery linked to 2 night Inpatient | + (0) | X | X | + (2) | X | + (9) | + (95) | X | 106 |
| Emergency Room With Surgery linked to 3+ night Inpatient | + (0) | X | X | X | X | + (0) | + (2) | X | 2 |
| Emergency Room No Surgery linked to 3+ night Inpatient | + (2) | X | X | X | X | + (13) | + (267) | X | 282 |
| Other Emergency Room, With Surgery | + (3) | X | X | + (3) | + (8) | + (34) | + (115) | + (14) | 177 |
| Other Emergency Room, No Surgery | + (8) | X | X | + (9) | + (122) | + (31) | + (1426) | + (144) | 1,740 |
| Medical Visit | X | X | X | + (67) | + (322) | + (3) | + (23) | + (22,787) | 23,202 |
| Total | 260 | 207 | 605 | 1,027 | 10,153 | 116 | 2,314 | 32,966 | 47,648 |

ER = emergency room; HC = Household Component; HS = hospital stay; MPC = Medical Provider Component; MV = medical visit; OP = outpatient.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.  
Note: + = match allowed, X = match not allowed. The number of events that matched is given in parentheses.

Exhibit 8.8 MPC-HC Matches Allowed and Not Allowed at Pass 1.5, 2022

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MPC Classification** | **HC Classification** | | | | | | | | |
| **HS, 0, 1 Night** | **HS, 2 Nights** | **HS, 3 Nights** | **OP, With Surgery** | **OP, No Surgery** | **ER, With Surgery** | **ER, No Surgery** | **MV** | **Total** |
| Hospital Stay, 0, 1 Night | + (0) | + (0) | + (1) | X | X | X | X | X | 1 |
| Hospital Stay, 2 Nights | + (0) | + (3) | + (1) | X | X | X | X | X | 4 |
| Hospital Stay, 3 Nights | + (1) | + (2) | + (14) | X | X | X | X | X | 17 |
| Outpatient, With Surgery, multi-night observation not indicated | X | X | X | + (25) | X | + (3) | X | X | 28 |
| Outpatient, No Surgery, multi-night observation not indicated | X | X | X | X | + (1,733) | X | + (54) | + (1,800) | 3,587 |
| Outpatient, With Surgery, multi-night observation indicated | X | X | X | + (13) | X | + (1) | X | X | 14 |
| Outpatient, No Surgery, multi-night observation indicated | X | X | X | X | + (14) | X | + (1) | + (46) | 61 |
| Emergency Room With Surgery linked to <2 night Inpatient | X | X | X | + (0) | X | + (0) | X | X | 0 |
| Emergency Room No Surgery linked to <2 night Inpatient | + (0) | X | X | X | + (1) | X | + (8) | X | 9 |
| Emergency Room With Surgery linked to 2 night Inpatient | X | X | X | + (0) | X | + (0) | X | X | 0 |
| Emergency Room No Surgery linked to 2 night Inpatient | + (0) | X | X | X | X | X | + (6) | X | 6 |
| Emergency Room With Surgery linked to 3+ night Inpatient | X | X | X | X | X | + (0) | X | X | 0 |
| Emergency Room No Surgery linked to 3+ night Inpatient | + (0) | X | X | X | X | X | + (10) | X | 10 |
| Other Emergency Room, With Surgery | X | X | X | + (3) | X | + (1) | X | X | 4 |
| Other Emergency Room, No Surgery | + (0) | X | X | X | + (133) | X | + (67) | + (144) | 344 |
| Medical Visit | X | X | X | X | + (217) | X | + (34) | + (3,992) | 4,243 |
| Total | 1 | 5 | 16 | 41 | 2,098 | 5 | 180 | 5,982 | 8,328 |

ER = emergency room; HC = Household Component; HS = hospital stay; MPC = Medical Provider Component; MV = medical visit; OP = outpatient.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

Note: + = match allowed, X = match not allowed. In addition, inpatient events were required to have a difference in length of stay less than 3 days to match. The number of events that matched is given in parentheses.

**Comment 5: FORCE1 Variable**

The DONT# variables discussed above are used in the matching software so that the software will not pair up the unallowable events. To accomplish the opposite—in other words to force a connection between a particular HC and MPC event—the FORCE1 variable can be used. When an HC event and an MPC event have the same value for FORCE1 within a block group, the matching algorithm assigns the pair an extremely large match field weight, specifically an agreement weight of 10,000. The FORCE1 variable is rarely used and only after a hand review of selected matched pairs reveals some unappealing matches are being made within block groups.

# 2022 Medical File Matching Process—Quality Control Checks

Given the importance of the matching algorithm to the Medical Expenditure Panel Survey (MEPS), RTI conducted a series of extensive quality control (QC) checks to verify that the algorithm is working as expected and that the results are accurate. The QC processes include the following steps:

* RTI verifies within each block group that there are no instances where events exist on both the Household Component (HC) and Medical Provider Component (MPC) files that are not linked, unless these matches are not allowed (see **Exhibit 8.7)**. As noted previously, one of the enhancements that RTI made to the process beginning with the 2009 MPC was to link as many events as could be linked within a block group regardless of the size of the match weight. RTI also verifies that no unallowable matches are being made as noted in **Exhibit 8.7**.
* RTI compares distributions of the variables used in matching for the matched results between the most recent 3-year period.
* RTI identifies unmatched MPC or unmatched HC Hospital events. All events that are allowed to match with a hospital event for each person who has an unmatched hospital event are examined. This review may identify providers in two different collapsed contact groups for the same patient that appear to be the same and are not being caught during the process of creating the collapsed contact group identifier (ID) number. In these cases, the two collapsed contact groups under review are combined by hand.

RTI also examines cases in which unmatched HC and MPC events occur for a person on the same date. If there are providers in two different collapsed contact groups that appear to be the same and are not being caught during the process of creating the collapsed contact group ID number, the two collapsed contact groups under review are combined.

Another possible review is of somewhat undesirable matches. Specifically, reviews can be conducted of inpatient stays and ER events whose event begin date differs significantly from the begin date of the event with which it is matched. Similarly, another possible review is an examination of paired inpatient stays with a very large difference in HSPLUS1and pairs with certain differences in EVENT. If a pair is flagged for review via the QC process, all HC and MPC events for the patient are placed into a spreadsheet so it can be determined if other matching scenarios make more sense for the patient (other than the one being made by the matching program). None of the pairs that were matched were changed because of this review in 2019, although the review did suggest some changes to the collapsed contact grouping variable (GID2). This review usually results in very few instances where collapsed contact groups need to be combined and was not conducted in 2021. In general, the approach is not to break any matches or force other matches unless the evidence clearly indicates that pairs should not have been matched or should be matched.

In addition, during the development of the algorithm, RTI verified that it performed as intended. A sample of MPC-HC pairs was examined, and the computation of the match weight was verified by saving all matching variables (fields) to a spreadsheet and setting up formulas that replicated the field comparison score of each variable. The computation of the *m* and *u* probabilities used in the comparison score for all fields was then verified.

# 2022 MPC Home Health and Home Non-Health Data

This section discusses the editing and matching associated with the Home Health (HH) and Home Non- Health (NHH) data. Several acronyms are used to describe these data files. The HH files are also referred to as the Home Care Health (HCH) files; the NHH files are also referred to as the Home Care Non-Health (HCN) files.

## Preparing the HH and NHH Data for Editing

Prior to editing of the HH and NHH data, the two data files are combined into one file. As depicted earlier, after editing is complete and prior to matching, the combined, edited file is broken apart into the edited versions of the HH and NHH data files.

## Editing HH and NHH Data

The steps in the editing and checking of the HH and NHH data mirror those for the Medical File data. The process of inserting missing codes to reflect skip patterns and other types of nonresponse, recoding variables in preparation for matching, and creating new variables is repeated for the HH and NHH data files. This section discusses issues unique to the HH and NHH data.

In preparation for matching, the HH and NHH data are summarized to the month level by the Medical Provider Component (MPC) person-by-provider pair ID (PPID). The raw Blaise files contain events with billing periods that span various time periods (month, 60-day period, or other). To merge with the Household Component (HC) data, the MPC data are first converted to the PPID by month level. An event that spans multiple months is broken into multiple events with payments and charges prorated according to the proportion of the billing period that occurs in each month. If multiple events occur within a month for a person–provider pair, these events are summarized into one event for the pair. In the remainder of this section, a HH “event” is the month by PPID summary level of reporting.

Prior to matching, the lowest charge events were excluded from matching and the charges and payments from the low-charge events were allocated to the remaining events. This change was made at Agency for Healthcare Research and Quality’s (AHRQ’s) request. A comparison was made between the number of MPC and HC month by PPID events for each person. When the number of MPC events was larger than the number HC events, the MPC events with the lowest (including zero or missing total charge) were identified. For each person where the number of MPC events exceeded the number of HC events, up to three of the low-charge events were identified and the charge and payment data in those low-charge events were distributed equally across the remaining MPC events. The low-charge events were then excluded from matching. In other words, if there were one, two, or three more MPC events reported than HC events, then one, two, or three lowest charge MPC events were excluded from matching and their charges and payments distributed over the remaining event. If there was a difference of more than three events, then only the three lowest charge events were selected to be excluded.

Sometimes there are more MPC months than HC months and the first calendar MPC month may include charges for a small number of days. The matching is by rank order rather than actual month. This can result in MPC months with a small number of days and small charges being paired with a full month from the HC file and some matched events having relatively small total charges for the services provided. Because of this, AHRQ modified the matching specifications after the 2015 MPC so that when there are more MPC months than HC months, the events with low-charge months are excluded from matching. The original charge and payment variables are included on the files in the variable names that begin with “ORIG\_.” For example, TOTLCHRG contains the total charge for the event after allocating the low-charge months and ORIG\_TOTLCHRG contains the original total charge. Similarly, TOTLPAYM and ORIG\_TOTLPAYM contain the total payment after the allocation and the original total payment for the event.

## Matching of the HH Data

After excluding the low-charge events, the 2022 HH matching used the same procedures that have been in place since the 2011 MPC. The matching is done in rank order: first HC month to first MPC month, etc. For example, if the HC has May-June-July and the MPC has July-August-September, matches are made as shown below. The months were rank ordered on both sides (1 to 3 in this example) and then merged. The only exception to this procedure is when the number of MPC events was larger than the number of HC events for a person, then up to three of the lowest charge MPC events were excluded from the matching.

|  |  |  |  |
| --- | --- | --- | --- |
| **HC** | **Rank Order** | **MPC** | **Rank Order** |
| May | 1 | July | 1 |
| June | 2 | August | 2 |
| July | 3 | September | 3 |

HC = Household Component; MPC = Medical Provider Component.

Source: 2022 Medical Expenditure Panel Survey – Medical Provider Component.

Two passes are used for matching the files. Like the matching of the Medical Files, at the first pass, the matching is within person–provider pairs using the rank ordering illustrated above. At the second pass, any unmatched events from the first pass are matched within the person (across all providers). Most person–provider pairs differ in the number of HC and MPC months (the length of the spell). At the second pass, using the nonmatches from the first pass, the months are again rank ordered before matching across providers. For example, if in this case the HC does not include July, the September record in the MPC is rank ordered again along with the person’s other residuals before matching to any HC residuals (also freshly rank ordered).

**Exhibit 11.1** summarizes the HH matching process for the most recent 3-year period. The number of person–provider pairs on the HC and MPC have remained stable for these 3 years like the number of events on the files.

Exhibit 11.1 Summary of HH Matching Process, 2020–2022

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2020 | 2021 | 2022 |
| Number of Person–Provider Pairs on HC File | 593 | 666 | 604 |
| Number of Person–Provider Pairs on MPC File | 531 | 584 | 528 |
| Number of HC Events (months) on File | 3,119 | 3,651 | 3,078 |
| Number of MPC Events (months) on File | 2,527 | 2,832 | 2,356 |
| Number of Pass 1 Matches | 2,360 | 2,629 | 2,176 |
| Number of Pass 2 Matches | 73 | 98 | 90 |

HC = Household Component; HH = Home Health; MPC = Medical Provider Component.

Source: 2020–2022 Medical Expenditure Panel Survey – Medical Provider Component.

After the HH data are matched, condition codes for matched pairs are examined as a QC check of the matching. ICD-10, CCSR, and CCSRMATCH codes from the HC and MPC files are compared for matched pairs. (The CCSRMATCH codes are recoded versions of the CCSR variables. They are used in the OBD/Hospital matching, and the same recode is used on the HH files for evaluation). The number and percentage of matched pairs that have at least one code that agree. The interpretation of conditions by the household respondent may differ from that of the provider. The CCSRMATCH codes are recodes of the CCSR codes and are more general than the CCSR codes. Hence, they are less susceptible to nonsampling type error, and more matched pairs that agree with respect to the CCSRMATCH codes are expected.

## Summary of Key 2022 MPC HH and NHH Variables

Key variables from the HH and NHH files are the payment and charge variables, and variables that give the reason for discrepancy between the payments and charges. RTI prepares summary tables that examine these variables across the most recent 3-year period, by type of event (HH or NHH), and source of payment. These distributions and summary statistics are compared to identify anomalies or outliers in the data.

# 2022 MPC File Delivery and Preparing Codebooks

To provide documentation and facilitate use for AHRQ analysts and other interested researchers, RTI produces and provides a Microsoft Word formatted codebook for each edited file delivered. Each codebook has the same structure and contains the following information:

* A summary of the contents of each data file (e.g., total numbers of variables, total number of records)
* A table of contents with hyperlinks to the sections that provided variable-level information
* A listing of every variable on the file in alphabetical order and by order of appearance; these indexes include hyperlinks to the more detailed information on each variable
* A detailed description of each variable, including the variable label, the type of variable (numeric or character), a list of values and the associated value labels, and weighted and unweighted frequencies for categorical variables; for continuous variables, univariate statistics are provided in lieu of frequencies (e.g., mean, maximum, and minimum)
* The 2022 Event Form, which is used to record the information on an event), including corresponding questions and the variables associated with each question.

# Editing and Matching Procedures for the 2023 MPC

For the 2023 MPC editing and matching task, RTI plans to closely follow the process used for the 2022 MPC. The following sections discuss specific revisions and additions planned for the editing and matching of the 2023 MPC data. This list will be appended if additional changes are identified during small-group discussions of matching during Summer 2024.

## Plans for Editing

The data editing checks that will be used in the 2023 MPC are like those used for the 2022 MPC. Spreadsheets containing the events to be reviewed are produced at least weekly during data collection, and several times after data collection has ended, until all potential problems have been reviewed and resolved. Data collection managers review and resolve the events that are identified as meeting the various criteria. For the 2023 MPC, the reviews were started in February 2024, soon after the start of the MPC data collection.

The items listed below are suggested changes for the 2022 files based on changes in the Blaise data collection system, Edit Shop, and comments from AHRQ on the prior year files. As in the 2020 – 2022 cycles, VA pair data will not be a part of the 2023 MPC data collection and abstraction activities. The VA will provide AHRQ data for those pairs identified as VA providers/facilities. Data for these pairs will be assigned ID variables and values comparable to the MPC values (e.g., GID, PPID, and PMEID), and the data concatenated with the Blaise data collected for the other sampled pairs. Consistency codes for MPC variables that do not apply to the VA records will be the same as in the 2020 - 2022 cycles. Any changes or edits to the VA events so that the variables conform to the MPC variables will be documented in memoranda that accompany the data files that are delivered to AHRQ.

The master lists of providers were updated during the 2022 cycle. RTI will continue to examine providers that are not classified as one of the four types and contain “County” or “District” in the provider’s name and will classify those that are government-owned or government entities as “CHARITY” In addition to internet searches, RTI will use again the hospital ownership field of the Centers for Medicare and Medicaid Services Hospital General Information file (<https://data.cms.gov/provider-data/dataset/xubh-q36u>) and will verify this information using internet searches. RTI suggests appending military bases that were renamed in 2023 to the master list of military providers.

1. Contact groups are clusters of providers that share the same point of contact. This information is used for data collection purposes and later during matching of HC and MPC events. An appropriate point of contact is identified and asked to respond for all patients and providers associated with a particular contact group. The contact group identifier is later used as a blocking variable in matching. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. The GID is constructed during the MPC sample preparation but may change during data collection. For example, if two separate GIDs have the same POC, they may be combined or merged. The GID often represents one provider or health care system, but may contain providers from multiple systems or group practices when the contact has access to the records from other providers.. [↑](#footnote-ref-3)