

---

# Updated Master Patient Index and Data Exclusion

## Summary

The Center for Health Information and Analysis (CHIA) created the Master Patient Index (MPI) process to track data associated with specific individuals without using Personally Identifiable Information (PII). This allows CHIA to provide data that has research utility for studying individuals over time but does not have enough identifying information to pose a significant privacy risk.

In the MA APCD, CHIA provides a derived data element called Member Link EID, also called the MEID (`memberlinkid`). Member Link EID is a surrogate key, which is a unique reference that does not contain identifying information. Member Link EID represents an individual across Member Eligibility and Claims files and across different payers/carriers. This enables analysis of an individual's data even as their insurance coverage changes over time.

In the CY 2024 release, Member Link EID appears in a separate MEID file that includes Member Link EID, Submitter/Org ID (`orgid`) and CHIA Carrier-Specific Unique Member ID/CSUMID (`chiacarrierspecificuniquememberid`). Using `memberlinkid`, you can find all Org IDs and CSUMID records associated with the same individual. You can then search for each associated combination of `orgid` and `chiacarrierspecificuniquememberid` in the Member Eligibility and Dental, Medical, and Pharmacy Claims files to find data rows associated with this individual.

To use `memberlinkid` with the CY 2024 APCD release, use the CY 2024 MEID file. For guidance on using `memberlinkid` in previous years' data, consult previous years' documentation or contact CHIA for assistance.

## MPI Process

Before payers submit data to CHIA, they process it using CHIA's secure file intake software. This software performs data quality checks, standardizes formatting of certain fields, and removes invalid characters. The software also "hashes" plain-text personal identifiers, such as name and Social Security number, by replacing them with randomly generated 128-character references that do not contain sensitive information. Data submissions are then encrypted and sent through a secure portal. In this process, CHIA never receives plain-text PII.

When CHIA receives hashed data, it carefully reviews each submission for anomalies or duplications that might meaningfully decrease the utility of the MPI. Data submissions might be excluded, in whole or in part, for a few different reasons. For more information, see "Data Excluded from MPI Process" on page 2.

In the MPI process, CHIA trusts that submitting payers have assigned unique membership identifiers to each individual they insure. CHIA's hashed version of this identifier is the CSUMID. CHIA also assigns a unique organization identifier or Org ID to each data submitter. In Eligibility and Claims files, CHIA treats any records with the same combination of Org ID and CSUMID as information about the same person.

CHIA uses master data management (MDM) techniques to assign a Member Link EID to Eligibility and Claims records. These MDM techniques make probabilistic matches based on member characteristics, determining which records most likely belong to the same person. The records between insurance carriers do not have to be completely identical in order to match, but the more similar that two records are, the more likely it is that they will match and receive the same Member Link EID.

For example, the records below have the same Member Link EID because the hashed versions of their name, month and year of birth, gender, and Social Security numbers are the same, even though their insurance information and ZIP Codes are different.

Member Link EID	Org ID	CHIA Carrier-Specific Unique Member ID	Hashed First Name	Hashed Last Name	Hashed Month/Year of Birth	Gender	Hashed SSN	ZIP Code
112233	22	HVD00000122	Hashed1	Hashed2	Hashed3	M	Hashed4	02118
112233	22	HVD00000122	Hashed1	Hashed2	Hashed3	M	Hashed4	02118
112233	30	BBY000034234	Hashed1	Hashed2	Hashed3	M	Hashed4	01056
112233	30	BBY000034234	Hashed1	Hashed2	Hashed3	M	Hashed4	01247
112233	30	BBY000034234	Hashed1	Hashed2	Hashed3	M	Hashed4	01702

These example records are simplified for clarity, but the actual Member Link EID element (`memberlinkid`) has type `Integer` and length 10.

CHIA continues to improve its MDM techniques to increase the utility of the MPI and decrease the number of “orphan” records, i.e. records with no match. Continuous improvement of the MPI process ensures that CHIA maintains security and patient privacy protections while delivering a robust, unique individual identifier that improves researchers’ ability to track members across time, sites of service, products, and payers.

## Data Excluded from MPI Process

### Completely Excluded Submitters

CHIA excluded the following submitters (listed by `orgid`) from the MPI process. Complete exclusions can occur for two reasons:

- Data from a third-party administrator (such as a pharmacy benefit manager) has many records that are duplicates of data from a payer. In this case, CHIA excludes data from the third-party administrator’s `orgid`.

- Data quality checks find that multiple unique individuals seem to have the same `orgid` and Carrier-Specific Unique Member ID (`chiacarrierspecificuniquememberid`). In this case, CHIA excludes data from this `orgid`.

Org ID
10942
10953
10954
11069
11239
11264
11364
11371
11500

### Year-by-Year Exclusion of Submitters

CHIA analyzes data from submitters by year (first four digits of `submissionyearmonth`) to determine how often multiple unique individuals seem to share the same combination of `orgid` and `chiacarrierspecificuniquememberid`. If certain years have many more of these cases than other years, CHIA removes data from these particular years rather than excluding a submitter's entire data set.

Every year, CHIA works with submitters to obtain the most accurate possible data. Several of these submitters improved their data quality and are included in the MPI process in more recent years.

Org ID	Excluded Years/Values
299	2011–2016
7249	2011–2018
7273	2015–2016
10928	2011–2017
10933	2011–2022
10934	2013–2018
11065	2013–2014, 2017–2022
11274	2012–2018
11426	2017–2018

## Record-Level Exclusion of Submitters

Occasionally, CHIA excludes specific records from the MPI process, rather than entire years or submitters as a whole. CHIA formed the list below by analyzing submitters across all years and determining how often multiple members were submitted with the same Org ID and CSUMID combination in the same year. In these cases, only a few combinations were used for multiple individuals.

Org ID	Excluded Years/Values
10920	Specific chiacarrierspecificuniquememberid values from source data
11237	6 instances of chiacarrierspecificuniquememberid from 201712 and 201806



CENTER FOR HEALTH INFORMATION AND ANALYSIS

501 Boylston Street, Boston, MA 02116

617.701.8100 | [www.chiamass.gov](http://www.chiamass.gov)