

# CHIA's New MA APCD Master Patient Index

## Background

In 2017, CHIA embarked on a series of changes to the way it collects Personally Identifiable Information (PII) for the MA APCD. CHIA understands the privacy concerns around holding this data and recognizes that reducing CHIA's patient identifying information footprint is in everyone's best interest. The changes include the obfuscation and in some cases removal of certain patient identifiers at the site of the insurance carrier. They additionally include the obfuscation and in some cases removal of PII housed at CHIA for both current MA APCD data and previous releases.

As part of the implementation of these changes, CHIA has created a new patient identifier that allows analysts to link patients within and across carriers.

## Summary

CHIA has created a new MA APCD Master Patient Index (MPI) that assigns a single unique surrogate key to each person, regardless of how many different insurance carriers have submitted data about the person. The MPI is created using Master Data Management (MDM) techniques by employing an industry-leading MDM software solution, IBM Initiate Master Data Service. CHIA has designed and implemented a MA APCD-specific record linkage algorithm within the IBM Initiate software that operates on obfuscated PII. This record linkage algorithm is used to create and maintain CHIA's MA APCD MPI.

## Input Data Preparation

Data obfuscation begins by processing Member Eligibility (ME) data through CHIA's data intake application which is deployed as on-premise software at the data submitter. This application performs data quality and transformation steps prior to masking PII. These steps allowed for significant improvement in using masked data as part of the MPI matching process. This stage includes:

1. Performing data quality checks on submitted fields.
2. Standardizing first names using USPS nickname handling to improve record linkage.
3. Standardizing first and last name such as:
  - a. Removing non-alpha characters.
  - b. Removing unnecessary whitespace.
4. Applying phonetic encoding algorithms to name fields to assist in record linkage.

In summary, the data that is transmitted to CHIA no longer contains plain-text PII. Names, dates, and identifiers are converted to hashed equivalents. In addition, CHIA no longer collects redundant PII fields in the claim data submissions.

All legacy MA APCD data submitted to CHIA prior to this deployment has been prepared and securely hashed using the exact same logic.

## Input File Analysis

CHIA employs a substantial analysis step, carefully reviewing all eligibility files for anomalies or duplications that would meaningfully decrease the utility of the MPI. As MA APCD consists of submissions from both the risk holder (payer) and, in some cases, an administrator, it was determined in some circumstances this created a significant amount of duplication. In these cases the administrator data is excluded from the MPI process. In other cases some submitters' data may be found to be of poor quality and be excluded. Each year new files will be reviewed for anomalies or duplications.

Documentation on data or files excluded from the MPI process can be found on CHIA's website at: <http://www.chiamass.gov/assets/docs/p/apcd/MA-APCD-Release-6.0-MPI-Data-Exclusion-Overview.pdf>.

## Record Matching

CHIA's MPI solution utilizes a probabilistic approach that uses select fields to generate a score that represents how well a record matches to another record. A high score denotes the records are considered a match and are linked together as a single entity (person). Records linked together are assigned a surrogate key known as a Member Enterprise ID, or MEID for short. The MEID in MA APCD is the MEMBERLINKID field.

## MEID Assignment to Claims

CHIA trusts that the insurance carriers that submit data to the MA APCD assign unique membership IDs to each individual insured by them. CHIA refers to these IDs as carrier-specific unique member IDs (CSUMIDs). In addition, CHIA assigns a unique organization ID to each data submitter (OrgID). CHIA treats any records with the same OrgID and CSUMID combination in either member eligibility or claim data as information about the same person.

The MEID surrogate keys are assigned to medical, pharmacy, and dental claim records based on their OrgID and CSUMID.

## Conclusion

After CHIA tested the newly created unique membership IDs, CHIA found significant improvement in the match rates and significant reduction in the number of "orphan" records. "Orphan" records are records with no connection to any claims or any other membership records.

This updated process ensures CHIA can maintain security and provide improved patient privacy protections while delivering a better, more robust unique membership ID that is homogenous over five years and improves the researchers' ability to track members across time, sites of service, products, and payers.

Note: Since values for unique membership IDs are recalculated prior to each new release, this surrogate key cannot be used to link data across any other prior or subsequent releases. CHIA is working on a surrogate key mapping across releases.