MHDO Analytic Dataset Proposal

Last Updated 1/11/2019

Contents

[Overview 1](#_Toc534970349)

[Current Structure 1](#_Toc534970350)

[Proposed Structure 2](#_Toc534970351)

[Medical Claims Data Structures 3](#_Toc534970352)

[Dental Claims Data Structures 3](#_Toc534970353)

[Pharmacy Claims Data Structures 3](#_Toc534970354)

[Member Eligibility Data Structures 3](#_Toc534970355)

[Provider Data Structures 4](#_Toc534970356)

[Appendix A Proposed Medical Claims Release Structures 5](#_Toc534970357)

[Medical Claims Header 5](#_Toc534970358)

[Medical Claim Line 7](#_Toc534970359)

[Provider File 8](#_Toc534970360)

[Medical Claim ICD Diagnosis Codes 9](#_Toc534970361)

[Medical Claim ICD Procedure Codes 9](#_Toc534970362)

# Overview

The new MHDO Data Delivery Model calls for the creation of “analysis ready” datasets that will allow data users to more easily analyze the data without additional restructuring or cleaning. This document outlines these new analysis-ready release structures and the steps that will be taken to create them.

# Current Structure

Currently, data from the MHDO APCD are released to data users as monolithic data tables accompanied by a series of support tables. The six main tables are for:

* Dental claim lines,
* Medical claim lines,
* Pharmacy claim lines,
* Dental eligibility,
* Medical eligibility, and
* Pharmacy eligibility.

The support tables are:

* Medical claims consolidation (identifies medical claim lines to use to obtain latest version of a given claim),
* Counties,
* Home-grown CPT codes,
* Home-grown diagnosis codes,
* Provider detail and master tables (one set each for dental, medical, and pharmacy claims),
* Payer codes,
* Provider specialty codes.

These formats can be difficult for some data users to make use of. For instance, in order to perform claim-level analyses, data users have to roll up individual claim lines into a claim-level record. Information about a claim’s ICD procedure codes is spread out across 25 different ICD-10 procedure code fields and one ICD-9 field. Information about a claim’s diagnosis codes is spread out across 53 ICD-10 fields and 15 ICD-9 fields. This structure can make even very basic queries quite complex.

Member eligibility records provide member demographic information such as gender, date of birth, and patient location on each monthly record. If there are variations in this information across time, this can lead to analytic problems when reporting by age or patient location.

# Proposed Structure

The proposed analytic structures will provide simpler, summarized data structures that data users will be able to take advantage of. Information that is unique to a claim or member will be cleaned or removed from the claim-level or member-level tables with a flag to indicate invalid data were removed. For example, a diagnosis code not on the ICD-10 list will be removed. The original value would be present in the base data as currently available. In addition to dropping invalid values we will also drop non-standard “home-grown” codes. In general, these codes are not useful for standard reporting and value-add processing since they are only specific to a given payer and then only for a period of time, typically. Fields containing values that are valid but malformed will reformatted rather than being removed and flagged. So, for instance, a revenue code value of “300” would be reformatted as “0300” in the analytic file.

When there are several different fields conveying similar information, such as diagnosis code information, this information will be transposed into a narrow table that can be more easily queried. For example, the data submission layout allows for the submission of 24 ICD-10 “Other Procedure” Codes. Currently, if analysts are interested in querying claims that include a certain procedure code, they have to include each of these 24 fields in their query. The narrow table structure (see Appendix A for an example) will have each of the 24 fields as its own row (assuming all fields were populated). Thus, a query only needs to reference a single field on this table to look across all procedure codes. Code type, sequence number, and ICD Version fields will provide contextual information about each code, allowing, for instance, the isolation of principal procedure codes.

Provider information will be simplified so that each claim is unambiguously associated with a single record on a provider composite table that covers all provider types (a claim can have servicing/rendering, billing, attending, operating, and referring providers listed). The same provider table will be used for dental, medical, and pharmacy claims.

Critical data quality issues on records will be communicated to end users using flag variables; this will allow data users to easily include and exclude records based on the scope of their analysis. The use of flag variables will eliminate the need to distribute a separate claim consolidation table and provide the most flexibility to end users.

Finally, we will be removing the data submission field indentifiers (MCxxx, MExxx etc.) which currently prefix many variable names in favor of more user-friendly naming. We will also use this as an opportunity to clean up existing abbreviations such as “REL” to “Relationship Code” to be clearer. These “user-friendly” field names will be fully documented in the data dictionary, allowing data users to easily determine each field’s derivation.

## Medical Claims Data Structures

The medical claims will be restructured into the following tables (see Appendix A for details):

* Medical Claims Header—one row per claim providing all claim-level information
* Medical Claim Line—one row per claim line with all claim line-level information
* Medical Claim DX—one row for each diagnosis code on claim.
* Medical Claim PX—one row for each ICD procedure code on claim (Please note: HCPCS/CPT procedure codes would be present in the Medical Claim Line table only.)

Note that claim DX and ICD-based PX codes are expected to represent claim-level information. However, the data cleaning process will test for situations where the codes differ across lines of the same claim and ensure that codes appearing on any claim line are captured; the claim would then be flagged as having a data quality issue.

## Dental Claims Data Structures

The Dental claims will be restructured into the following tables:

* Dental Claims Header—one row per claim providing all claim-level information
* Dental Claim Line—one row per claim line with all claim line-level information

## Pharmacy Claims Data Structures

The Pharmacy claims will be restructured into the following tables:

* Pharmacy Claims Header—one row per claim providing all claim-level information
* Pharmacy Claim Line—one row per claim line with all claim line-level information
* Pharmacy—one row for each pharmacy appearing in a claim

## Member Eligibility Data Structures

In the current data structures, dental, medical, and pharmacy eligibility data is stored on separate tables. However, this information comes to us all on a single table. Each record has a set of flags indicating what type(s) of eligibility it is for. Thus, a single source record could end up being stored in three different places. For the analysis-ready datasets, these records will be reunified. Thus, we will only have one set of eligibility data structures. However, data requesters will still only obtain access to the type of eligibility data they request (dental, medical, pharmacy, or a combination of these types).

An additional table is being considered to aggregate eligibility information. It would summarize for each member all of the types of eligibility it had in a given month, across all payers.

Eligibility table example

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| mhdo\_member\_id | year | month | com\_flag | mcr\_flag | mcd\_flag | dental\_flag | pharm\_flag |
| 100 | 2016 | 1 | 1 | 1 | 0 | 0 | 1 |
| 100 | 2016 | 2 | 1 | 1 | 0 | 0 | 1 |
| 100 | 2016 | 3 | 1 | 1 | 0 | 0 | 1 |
| 100 | 2016 | 4 | 1 | 1 | 0 | 0 | 1 |
| 100 | 2016 | 5 | 1 | 1 | 0 | 0 | 1 |

Please note that the claim headers include member demographic information. If we were guaranteed that every claim would be able to be associated with an eligibility record, we could omit these fields. However, since we know we don’t have 100% linkage between claims and eligibility in the MHDO APCD, we propose that we continue to supply this information on the claim records. When this information conflicts with the demographics provided in the eligibility data, the claim record will be flagged.

The MHDO-Assigned Member ID is constructed from the date of birth and gender information provided by the payer, along with either a member SSN, a subscriber SSN, or a contract number. When it is constructed from a contract number, it only identifies a member within a particular payer contract relationship. When it is constructed from an SSN, it generally can be used to identify a person between payers that also provide an SSN. The MHDO-Assigned Member ID will be revised to include additional identifiers (HICN and MBI) once we start receiving these fields in early 2019.

The member eligibility will be restructured into the following tables:

* Member—one row per MHDO-assigned member ID providing demographic information such as gender and date of birth. Note that information on member’s location of residence will not appear on this table, since this may change over time.
* Member Eligibility—one row per valid member eligibility record (after unification of dental, medical, and pharmacy into one record). This will generally be one row per member per month per payer. However, there may be cases where a member has two or more types of eligibility from a given payer in a given month (i.e., multiple insurance product types). Member location of residence will appear on this table since it is tied to a particular point in time, payer and insurance type.

## Provider Data Structures

* Provider—one row for each provider appearing in a claim, regardless of provider type (i.e., provider A will only have one row, even if it appears as both a rendering and a billing provider)

# Appendix A Proposed Medical Claims Release Structures

Note: Actual release structure column names are yet to be determined but will be unique to the release structures. Data quality flags are not shown, since they haven’t been finalized yet. As noted above, the final tables will have “user-friendly” column names rather than what is shown here. For instance, instead of MC003\_Product, the column would be named something like Insurance\_Product\_Type.

# Medical Claims Header





## Medical Claim Line



Note that the Provider IDs shown above will differ from the PRVIDN values currently used. The analytic files will use provider IDs that have a 1-to-1 mapping to the NPI of the entity. Thus, any time a particular provider is referenced on a claim line, it will have the same provider ID assigned.

## Provider File

The provider file will be based on information from the NPPES registry. There will be a 1-to-1 mapping between provider IDs and NPIs. The reason that we are not just using the provider ID here is to accommodate situation where the data user is not authorized to see provider identifiable information. In such cases, the data users would only receive non-identifiable fields such as provider specialty (PRVSPEC) and taxonomy code.

While the MHDO is not providing structural data on provider relationships at this time, it is exploring the feasability of providing this information. If this information is made available to analytic file data uses, it will make use of the same provider IDs that appear on the table below.



## Medical Claim ICD Diagnosis Codes

Note that the principal diagnosis code will also appear on the claim header, for convenience.

The DX\_Code\_Type will indicate the type of diagnosis code (admitting, e-code, reason for visit, principal, or other). The sequence number will allow the relative position of e-codes, reason for visit, and other DX codes to be determined—it will be null for principal and admitting diagnoses. The DX Code Type field will allow analysts to isolate rows for specific types of diagnoses, such as principal diagnosis codes of reasons for visit. The sequence number fields will allow analysts to determine the order that diagnosis codes of each type appeared on the claim (e.g., first e-code versus second e-code).



## Medical Claim ICD Procedure Codes

Note that the principal procedure code will also appear on the claim header, for convenience.

The PX\_Code\_Type will indicate the type of diagnosis code (principal or other). The sequence number will allow the relative position of other PX codes to be determined—it will be null for principal procedure. Note that HCPCS and CPT codes do not appear on this table. They will appear only on the claim line detail data structure. Just as with the diagnosis codes above, the code type and sequence number fields will allow for the isolation of principal procedure codes from other codes and the determination of the relative position of code appearance.

