

**MHDO’s Chapter 730 Data 2015-2021 – Summary of MHDO-assigned replacement Person ID Methodology**

Last updated: 11/30/2022

# Table of Contents

[Background 1](#_Toc120700713)

[Methodology 2](#_Toc120700714)

[Vital Statistics Birth Records 2](#_Toc120700715)

[Vital Statistics Death Records 4](#_Toc120700716)

[Cancer Registry Data 6](#_Toc120700717)

# Background

MHDO Chapter 730 Interagency Reporting of Cancer-Incidence Registry and Vital Statistics Data is a joint rule to provide for the reporting of cancer data and vital statistics data by the Department of Health and Human Services, Maine Center for Disease Control and Prevention; Data, Research, and Vital Statistics to the Maine Health Data Organization, pursuant to Public Law 2021, chapter 423.

Before 2020, MHDO’s person identification was based on a composite ID (version 1.0). This composite was based on DOB combined with either the member SSN, the subscriber SSN, or the contract number. Since SSNs aren’t available from many payers, this limited the ability to use person identifiers to link across data source because contract numbers are payer specific. MHDO hospital encounter data only included a composite based on encrypted MRN and, thus, this couldn’t be linked to the identifiers used in APCD data.

In 2020, MHDO revised its person identifier logic to take advantage new fields in its layouts to improve its ability to identify individuals with the MHDO-assigned replacement Person ID (version 2.0). This methodology a multi-level matching process that is more flexible and extensible than the older composite-based method. The availability of name and address in all data streams allowed for person identifiers to be assigned that allowed the linkage of hospital and APCD data.

Since the version 2.0 identification process can be adapted to account for different combinations of identifier fields, we were able to leverage this to assign person identifiers to the additional data streams described here. This will allow for the linkage of APCD, hospital, birth, death, and cancer registry data.

The assignment process outlined here assigns a MHDO-assigned replacement Person ID to each detail record similar to the APCD and hospital assignment processes. Each MHDO-assigned replacement Person ID links to information in MHDO centralized person index.

# Methodology

This section provides a description of the logic used to assign an MHDO De-Identified Replacement Person ID to the Chapter 730 Vital Statistics and Cancer Registry data. The Person ID will allow users to link Chapter 730 data with MHDO APCD and hospital encounter datasets. This section also provides summary level statistics regarding the matched and unmatched records.

## Vital Statistics Birth Records

Each birth record has up to three distinct individuals identified: the child, the mother, and the father. Since there may be correlation between these individuals (e.g., the same health plan or address), we identify the individuals in three stages. First, we identify the mother, then the father, then the child.

### Mother

To identify the mother, we create an extract of observed birth events from the APCD medical claims and hospital encounter data based on diagnosis and CPT codes. Our initial two passes attempt to match these birth observations with records from the birth data. After this, we performed matching against the entire person index, regardless of whether there was a birth observation in the claims or hospital data. Once a birth record was matched, it was eliminated from subsequent passes, ensuring each row received the best possible match.

APCD birth events were identified based on the following CPT codes: 59610, 59612, 59614, 59618, 59620, 59622, 59510 – 59515, 59400 – 59410. Past analyses have shown that these codes identify the majority of birth episodes that are pulled by diagnosis codes, so no additional diagnosis codes were used. For inpatient records, any diagnosis code in the range V30 – V39 or Z37 – Z38 was used to identify birth episodes. For outpatient records, both the CPT and diagnosis codes were used to identify birth events. Please note that the diagnosis code prefixes we look for include both ICD-9 and ICD-10 codes since our data contains a mixture of these. Some of the ICD-9 codes such as those associated with live birth (e.g., V30 Single liveborn) have different, non-birth-related meaning for ICD-10 (e.g., V30 Occupant of three-wheeled motor vehicle injured in collision with pedestrian or animal). Since our methodology uses other strong demographic factors in addition to diagnosis codes, this potential over-identification of birth events is highly unlikely to cause a misidentification since this would require the coincidence of many other factors, such as date of event, mother’s DOB, and name.

All name data were cleaned to remove non-space, non-alphanumeric characters and to use solely upper-case characters for matching purposes. Thus, strings such as “O’Brien” would qualify for a match to “OBRIEN”.

Since it is theoretically possible for one row to be matched to two (or more) different MHDO person IDs, code was added to choose the lowest person ID in these cases.

| Pass Number | Description |
| --- | --- |
| 1 | Match to birth observation with child date of birth (DOB) within event window on mother DOB, first name, and last name or maiden name |
| 2 | Match to birth observation with child DOB within event window on mother DOB, first four characters of first name, and last name or maiden name |
| 3 | Match to overall MHDO person index on mother’s DOB, first name, and last name |
| 4 | Match to overall MHDO person index on mother’s DOB, first four characters of first name, and last name or maiden name |
| 5 | Create new entry in MHDO person index and assign to birth record |
| 6 | Adjust existing assignments to choose lowest MHDO person ID when there are multiple matches for a mother |

There were 16 birth records lacking sufficient name or DOB information to identify the mother. **Overall, 95.6% of birth records had a mother who could be linked to an existing MHDO person index record (i.e., an APCD or hospital discharge record).**

### Match Rate by Pass

|  |  |  |
| --- | --- | --- |
| Match Pass | Count | Percentage |
|  | 16 | 0.0% |
| 1 | 51,190 | 61.3% |
| 2 | 448 | 0.5% |
| 3 | 27,823 | 33.3% |
| 4 | 155 | 0.2% |
| 5 | 3,215 | 3.9% |
| 6 | 626 | 0.7% |
| Total | **83,473** | **100.0%** |

### Father

While there is no predictable health event that we can use to help identify fathers, we can attempt to use the mother’s information to do so. Some fathers will share health coverage or an address with the mother. We will use these correlation observations for our initial rounds of matching before attempting to match to the overall MHDO person index. Cleaning steps and the other adjustments described for matching the mother are also performed for the father’s identifying characteristics.

|  |  |
| --- | --- |
| Pass Number | Description |
| 1 | Match to correlation observation on father DOB, father first name and father last name as well as one of the following with the mother: member or subscriber Social Security Number (SSN), member or subscriber name, address |
| 2 | Match to correlation observation on father DOB, father first name and father last name as well as a payer or submitter contract number with the mother |
| 3 | Match to overall MHDO person index on father’s DOB, first name, and last name |
| 4 | Create new entry in MHDO person index and assign to birth record |
| 5 | Adjust existing assignments to choose lowest MHDO person ID when there are multiple matches for a father |

There were 7,433 birth records with either no father or insufficient identifying information for the father to allow an assignment. **Overall, 84.7% of birth records have a father identified in the index; this percentage is 93.0% if only records with sufficient identifying information for the father are considered.**

Match Rate by Pass

| Match Pass | Count | Percentage |
| --- | --- | --- |
|  | 7,433 | 8.9% |
| 1 | 69,004 | 82.7% |
| 2 | 219 | 0.3% |
| 3 | 1,498 | 1.8% |
| 4 | 4,607 | 5.5% |
| 5 | 712 | 0.9% |
| Total | **83,473** | **100.0%** |

### Child

Similar to how we approached the identification of the father, we create an extract of corelated observations for each child based upon both the mother’s and the father’s information. The matching procedures then follow the procedure that was followed for the father.

The birth records appear to have the child’s first name padded out to 24 characters with spaces followed by the child’s middle initial in many, but not all, cases. To adjust for this and improve first name matching overall, child first name matching is either: the full first name, the first “node” of the name as delimited by spaces, or the first three nodes of the name as delimited by spaces (e.g., “Billy Jo Bob” would be parsed and matched into three separate name nodes).

|  |  |
| --- | --- |
| Pass Number | Description |
| 1 | Match to correlation observation on child DOB, child first name and child last name as well as one of the following with a parent: member or subscriber SSN, member or subscriber name, address |
| 2 | Match to correlation observation on child DOB, child first name and child last name as well as a payer or submitter contract number with a parent |
| 3 | Match to overall MHDO person index on child’s DOB, first name, and last name |
| 4 | Create new entry in MHDO person index and assign to birth record |
| 5 | Adjust existing assignments to choose lowest MHDO person ID when there are multiple matches for a child |

There were 42 child records with “\*\*\*\*” for the first name that could not be assigned person IDs. **Overall, 97.2% of child birth records were able to be matched to an existing person index record.**

Match Rate by Pass

|  |  |  |
| --- | --- | --- |
| Match Pass | Count | Percentage |
|  | 42 | 0.1% |
| 1 | 80,595 | 96.6% |
| 2 | 75 | 0.1% |
| 3 | 432 | 0.5% |
| 4 | 2,325 | 2.8% |
| 5 | 4 | 0.0% |
| Total | **83,473** | **100.0%** |

## Vital Statistics Death Records

To improve our matching, all records with an “Expired” discharge code were pulled from the APCD, hospital encounter IP or OP data and a death date imputed. This allowed for some high confidence matches, especially when combined with SSN. Our initial matching passes attempted to match these death observations with records from the death registry (passes 1 – 5). After this, we performed matching against the entire MHDO person index, regardless of whether there was a death observation in the APCD or hospital encounter data. Once a death record was matched, it was eliminated from subsequent passes, ensuring each row received the best possible match.

All name data were cleaned to remove non-alphanumeric characters and upper-case characters for matching purposes. Thus, strings such as “O’Brien” would qualify for a match to “OBRIEN”.

Since it is theoretically possible for one row to be matched to two (or more) different MHDO person IDs, code was added to choose the lowest person ID in these cases (this didn’t occur in the current data). Code was also added to arbitrarily choose one death event when one MHDO person ID mapped to more than one (this occurred 6 times in the current Death Records data). Future work may use these situations to identify potential splits or joins in the MHDO person index.

|  |  |
| --- | --- |
| Pass Number | Description |
| 1 | Match to death observation on date of death (DOD), DOB, SSN, first name, and last name |
| 2 | Match to death observation on DOD, DOB, SSN, and first four characters of first name |
| 3 | Match to death observation on DOD, DOB, and SSN |
| 4 | Match to death observation on DOD, DOB, first four characters of first name and last name |
| 5 | Match to death observation on unique combination of DOD and DOB in death registry |
| 6 | Match to overall index on DOB, SSN, First Name, Last Name |
| 7 | Match to overall index on DOB, SSN, first four characters of first name and last name |
| 8 | Create new entry in the MHDO person index and assign to death record |

**Overall, 75.1% of death records were able to be matched to an existing MHDO person index record (i.e., an APCD or hospital encounter record).**

Match Rate by Pass

| Match Pass | Count | Percentage |
| --- | --- | --- |
| 1 | 28,199 | 26.8% |
| 2 | 814 | 0.8% |
| 3 | 1,655 | 1.6% |
| 4 | 459 | 0.4% |
| 5 | 789 | 0.8% |
| 6 | 46,664 | 44.4% |
| 7 | 342 | 0.3% |
| 8 | 26,159 | 24.9% |
| Total | **105,081** | **100.0%** |

Because the person index generally only includes APCD assignments from 2017 on and hospital assignments from 2018 on the bulk of the 2015 and 2016 death records were not able to be matched in passes 1 – 5 (see table below). This is because, unlike birth events, decedents are unlikely to appear in claims or hospital data after their death dates.

Match Rate by Pass and Year

| Match Pass | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | Grand Total |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0.0% | 1.7% | 29.0% | 39.2% | 42.3% | 41.9% | 30.5% | 26.8% |
| 2 | 0.0% | 0.1% | 1.5% | 1.1% | 1.0% | 1.0% | 0.6% | 0.8% |
| 3 | 0.0% | 0.0% | 7.9% | 1.7% | 0.7% | 0.6% | 0.4% | 1.6% |
| 4 | 0.0% | 0.0% | 0.8% | 0.5% | 0.5% | 0.6% | 0.6% | 0.4% |
| 5 | 0.0% | 0.0% | 3.4% | 1.0% | 0.4% | 0.4% | 0.1% | 0.8% |
| 6 | 23.6% | 35.2% | 38.1% | 49.2% | 49.4% | 49.8% | 61.4% | 44.4% |
| 7 | 0.3% | 0.3% | 0.4% | 0.4% | 0.3% | 0.3% | 0.3% | 0.3% |
| 8 | 76.1% | 62.7% | 18.9% | 6.9% | 5.4% | 5.5% | 6.1% | 24.9% |

## Cancer Registry Data

We create an extract of observed cancer records from the medical claims and hospital data based on diagnosis codes. Our initial matching passes attempted to match these cancer observations with records from the cancer registry (passes 1 – 3). After this, we performed matching against the entire MHDO person index, regardless of whether there was a cancer observation in the MHDO APCD or hospital encounter data. Once a cancer record was matched, it was eliminated from subsequent passes, ensuring each row received the best possible match.

APCD and hospital records were considered to be associated with cancer events if any diagnosis code was in the range 140 – 209, 230 – 239, or began with ‘C’. This list was determined from a review of ICD-9 and ICD-10 code categories.

All name data were cleaned to remove non-alphanumeric characters and to use solely upper-case characters for matching purposes. Thus, strings such as “O’Brien” would qualify for a match to “OBRIEN”.

Since it is theoretically possible for one row to be matched to two (or more) different MHDO person IDs, code was added to choose the lowest MHDO person ID in these cases (there were fewer than 200 cases of ambiguous matches in these data).

| Pass Number | Description |
| --- | --- |
| 1 | Match to cancer observation on date of dx <= max observed cancer dx, DOB, SSN, first name, and last name |
| 2 | Match to cancer observation on date of dx <= max observed cancer dx, DOB, SSN, first four characters of first name |
| 3 | Match to cancer observation on date of dx <= max observed cancer dx, DOB, first four characters of first name, last name |
| 4 | Match to overall MHDO person index on DOB, SSN, first name, and last name |
| 5 | Match to overall MHDO person index on DOB, SSN, first four characters of first name, and last name |
| 6 | Create new entry in the MHDO person index and assign to cancer record |
| 7 | Match based on assignment of another cancer registry record with same CR0020\_Pat\_ID (note that this adjustment happens after pass 5 and again after pass 6 to reduce the number of new person IDs created). |

**Overall, 96.8% of cancer records were able to be matched to an existing MHDO person index record.**

Match Rate by Pass

|  |  |  |
| --- | --- | --- |
| Match Pass | Count | Percentage |
| 1 | 37,598 | 77.4% |
| 2 | 471 | 1.0% |
| 3 | 2,258 | 4.7% |
| 4 | 6,604 | 13.6% |
| 5 | 22 | 0.0% |
| 6 | 1,496 | 3.1% |
| 7 | 103 | 0.2% |
| Total | **48,552** | **100.0%** |