# New unique identifier methodology for APAC data

It is important for APAC to have an identifier that allows records to be correctly associated with a unique individual (deidentified) over time and across payers. Because Oregon APAC is prohibited from collecting Social Security numbers under ORS 442.373 (6), we cannot use Social Security numbers to definitively match records to individuals over time and through name changes. Instead, OHA and HSRI have worked together to determine the best method currently available to match data to unique individuals.

While this method effectively avoids assigning the same identifier to two different people, we believe it results in a single person being assigned more than one unique identifier in some cases, resulting in an overcount of unique people in APAC. At this time, we are opting to err on the side of missing matches rather than of potentially combining the records of different individuals incorrectly into one unique identifier.

The potential mismatches are largely related to the data that flows into APAC:

- To the extent address is used for matching, Oregon data is less likely to include a nine-digit zip code than more heavily populated states.
- To the extent city names are used for matching, Oregon has several instances of names that can be programmed over time but do not result in a quick/clean match. For example, Saint Helens, St. Helens, St Helens, Saint Helen, etc. HSRI reviews 'potential name problems' monthly for possible updates to the table used to normalize* the city names.
- Oregon allows flexibility in last name taken at marriage. Name changes at marriage or divorce may affect matching on member or subscriber last names, used in several steps of the algorithm.

OHA will continue to work with HSRI on incremental improvements to the matching process over time and will continue to be transparent with data requesters about this facet of our data. A description of the specific steps of the algorithm OHA agreed to use for matching appears below. Please contact us through apac.admin@dhsoha.state.or.us or discuss with the research analyst assigned to your approved project.

****************************

Description of the algorithm process from APAC's Business Rules document from vendor release 4 June 2021:

The Person ID is the unique identifier for a given person across all claims, payers, and years. When new eligibility and claims data submissions are received (data intake), attempts are made to assign these data to an existing Person ID from the Member Index. Then, an attempt to match data without an assigned Person ID is made in the data warehouse.

The vendor modified step 3 of the algorithm in release 4 in June 2021 and allowed Member Name (ME104 and ME105) to be 1 or more letters in the autojoin process.

The vendor modified the algorithm in release 4 in June 2021 and added a sixth step. The vendor used the nine-digit zip code reported by the Centers for Medicare and Medicaid services as the member address and modified the algorithm to substitute the nine-digit zip code. The nine-digit zip code identifies a geographic segment within the five-digit zip code i.e., city.

The vendor changed the auto join process release 4 in June 2021 to autojoin people who moved and changed payer plans in the same month.

Level 1 is a within payer match. Levels 2 to 5 are cross payer matches. the algorithm searches for existing combinations of fields (when populated and excluding default values) in the following levels in the order listed on incoming data only:

> 1. Payer Abbreviation (Code) and Member ID (ME010)
>
> 2. Subscriber Name (ME101 and ME102), Member Street Address (ME015A), Member City (ME015), Member State (ME016), DOB (ME014), Member First Name (ME105)
>
> 3. Member Name (ME104 and ME105), Member Street Address (ME015A), Member City (ME015), Member State (ME016), Member DOB (ME014)
>
> 4. Member Name (ME104 and ME105), Payer Code, Member ID (ME010) when length > 6, Member DOB (ME014)
>
> 5. Plan Specific Contract Number (ME009), Member ID (ME010), Member Name (ME104 and ME105), and Member DOB (ME014)
>
> 6. CMS nine-digit zip code (address) is matched to Member City (ME015), Member State (ME016) reported by other payers

Any individual that **does not match** an existing unique person identifier using the algorithm described above will appear in the APAC data as a new, unique person.

Notes on using the algorithm to create a unique identifier:
- DOB comparisons will only require a match on two of the three nodes (day, month, and year). Thus, minor typos in DOB shouldn't prevent a match.
- When member name or subscriber name are used (2 through 5 above), after basic cleaning of the name field, an exact match is required. However, the first name and last

name may be 'swapped' when performing the match. This means Henry Thomas will also be checked as Thomas Henry

- Wherever city and state matches are done, a match will also be performed on the five-digit ZIP code (ME017). A ZIP-based match will count as the city match.
- Wherever street address matches are done, a normalized address is used to make the match.
- APAC does not currently link records (assign a unique person identifier) solely using name and date of birth of an individual.

\* Normalizing data is a standard part of data cleaning. The purpose within the algorithm is increase the comparability of data 'pieces.' For APAC, this method is applied to names including persons, city, street, pre-directional and post-directional ('NE' treated as a match to 'Northeast'). This standardization improves the results of the algorithm.

*Once the APAC Data User Guide is available, we will add a link for more depth on this issue and potential impacts to use of the data.*