# A Possible Solution for Mapping OMOP Observation to PCORnet Vital Smoking

I have a solution for mapping smoking information from OMOP to PCORnet vital smoking fields. This solution meets one of my primary goals, as an ETL developer, that it can be implement as a single (though somewhat complex) SQL statement. The solution is complex, but if this were an easy problem we would not be still talking about it.

The first part the solution is to define a set of qualifying values that 1) can be filled from the various tobacco concept id’s and then be assigned to the PCORnet smoking variables.

Qualifying values

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Smoke** | **Smoke Status** | **Smoke Degree** | **Chew** | **Chew Status** |
| Yes | Current | Heavy | Yes | Current |
| No | X | Medium | No | X |
| Unknown | Never | Light | Unknown | Never |
|  | Unknown | Trivial |  | Unknown |
|  |  | Unknown |  |  |

**Smoke** – Indicates if the person smoked some tobacco product

**Smoke Status** – Indicates current status of the smoking use

**Smoke Degree** – Indicate the how much a person smoked

**Chew** – Indicates if a person uses/used non-smoked tobacco

**Chew Status** – Indicates current status of non-smoked tobacco

Starting with the various concept ids, it is possible to fill in the various qualifier values. For example, the concept id **266924008** (Ex-heavy cigarette smoker (20-39/day)) can be qualified as:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Smoke** | **Smoke Status** | **Smoke Degree** | **Chew** | **Chew Status** |
| Yes | X | Heavy | Unknown | Unknown |

And if this were all the information that we had we could fill in the PCORnet Vital table

|  |  |  |
| --- | --- | --- |
| **SMOKING** | **TOBACCO** | **TOBACCO\_TYPE** |
| 03=Former smoker | 03=Quit/former user | 05=Use of smoked tobacco but no information about non-smoked tobacco use |

The problem is a bit more complex, because there may be more than once tobacco related concept in OMOP. So that continuing with this example, the person may also have the following concepts.

**59978006** (Cigar smoker)

**228512004** (Never chewed tobacco)

Now filling out the qualifier values would look more like

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Concept Id** | **Smoke** | **Smoke Status** | **Smoke Degree** | **Chew** | **Chew Status** |
| **266924008** | Yes | X | Heavy | Unknown | Unknown |
| **59978006** | Yes | Current | Unknown | Unknown | Unknown |
| **228512004** | Unknown | Unknown | Unknown | No | Never |
| **Result** | Yes | Current | Heavy | No | Never |

And the values for PCORnet would now be

|  |  |  |
| --- | --- | --- |
| **SMOKING** | **TOBACCO** | **TOBACCO\_TYPE** |
| 01=Current every day smoker | 01=Current user | 02=Non-smoked tobacco only |

Note: This may not be the best fit. The person might be better represented by saying the smoke degree is *Unknown* since we don’t know how often they smoke a cigar which would result choosing a different value, *05=Smoker, current status unknown*, for **SMOKING.** (This is still a work in progress.)

When combining the qualifier values to get a single result it will be necessary to provide a ranking of values. For example, in the **Chew** column the result should be *No* so that *No* should rank higher than *Unknown.*

So the basics of this solution is

1. Qualify OMOP concepts into a set of values that provide sufficient information to populate PCORnet
2. Have an ordering for these values so that any number of rows can be reduced to a single row that best represents the answer for the possible qualifier values.
3. Provide a mapping for that single row of qualifier values to the three PCORnet tobacco related values.

Included is a work book with the following tabs:

* Variables – A table with the qualifier values above with an assigned rank.
* OMOP Concepts for Tobacco – List the concepts that we agreed upon for tobacco
* Concepts Quantified – Application of breaking the concepts into the qualifier values given above and then mapping that single concept to PCORnet (this is simplistic view because as shown above we need to assume multiple tobacco related concepts)
* Complete map – which is a mapping of all the possible combination of qualifiers mapped to PCORnet. The complete set is 3 possible values for **SMOKE**times 4 possible values for **SMOKE STATUS** times 5 possible values for **SMOKE Degree** and so forth for 3x4x5x3x4 equal 720 possible states