Cardio Harmonization Data Dictionary

## Conventions

* An artificial key is an “id”
* Vocabulary names are the names of the vocabularies: SNOMED, LOINC, etc
* Concept names are the concept identifiers within the vocabularies
* Concept ids are the ids given to concepts by OHDSI from a unified field of artificial ids so you don’t get id collisions between vocabularies.

## OHDSI CDM tables

* Person, keyed by person\_id  
  This table needs to be populated so that the foreign keys on person\_id are satisfied.
* Observation: fk concept\_id, person\_id, observation\_concept\_id
* Concept
* Measurement: fk measurement\_id, person\_id, measurement\_concept\_id
* Condition: fk condition\_id, person\_id, condition\_concept\_id
* visit\_occurence
* procedure\_occurrence
* death

## Harmonization Metadata Tables

* Study
* study\_id | study\_name | person\_id\_range\_start | person\_id\_range\_end
* ----------+------------+-----------------------+---------------------
* 1 | BEST       |                     0 |            10000000
* 2 | HFACTION   |              10000001 |            11000000
* 3 | SCDHEFT    |              11000001 |            12000000
* 4 | TOPCAT     |              12000001 |            13000000

Maps an integer study\_id to a study\_name. Also maps to a range of id values.

* + Study\_id
  + Study\_name
  + Person\_id\_range\_start
  + Person\_id\_range\_end
* Study\_to\_ohdsi\_mapping
* s\_id | from\_table |  f\_col  |   fun    | vocab | concept |  to\_table   |    to\_column
* ------+------------+---------+----------+-------+---------+-------------+-----------------
* 1 | best.lab2  | latri   | identity | LOINC | 2571-8  | measurement | value\_as\_number
* 1 | best.clab2 | clatrix | identity | LOINC | 12951-0 | measurement | value\_as\_number
* 1 | best.clab2 | clauax  | identity | LOINC | 3084-1  | measurement | value\_as\_number
* 1 | best.muga  | murve   | identity | LOINC | 10231-9 | measurement | value\_as\_number
* 1 | best.lab2  | laua    | identity | LOINC | 3084-1  | measurement | value\_as\_number

This is the main table for describing the mapping from a study’s tables to vocabularies and concepts within tables in the OHDSI CDM. The word “migration” is used to reference the process of moving data from a study to the CDM.

* + Study\_id: FK to study.study\_id
  + From\_table: the name of the table in the source study. Dotted with the schema name. Ex: Best.adju
  + From\_column: the name of a column within the source study table. Ex: id
  + Vocabulary\_name: the name of the vocabulary that contains the concept to which values from the named table, column will be mapped to. Ex: SNOMED, LOINC
  + Concept\_name: the concept ID used within the vocabulary. Ex. 3084-1
  + Function\_name: the name of a function used to transform the value. These appear in the module migrate\_functions.py. Ex: best\_sex\_to\_concept().
  + To\_table: the table in the OHDSI CDM where values will appear. Ex. Observation, Condition or Measurement
  + To\_column: the column within the table. Ex. Value\_as\_string, value\_as\_number, value\_as\_concept\_id
* Ohdsi\_calculation\_function
* s |       function\_name       |  vocab  |  concept  |  to\_table   |      to\_column
* ---+---------------------------+---------+-----------+-------------+---------------------
* 1 | best\_logical\_concept\_or   | SNOMED  | 6374002   | observation | value\_as\_concept\_id
* 1 | best\_logical\_concept\_or   | SNOMED  | 368009    | observation | value\_as\_concept\_id
* 1 | best\_logical\_concept\_or   | SNOMED  | 22298006  | observation | value\_as\_concept\_id
* 1 | sum                       | LOINC   | 71938-5   | observation | value\_as\_number
* 2 | concept\_or\_list           | SNOMED  | 53741008  | observation | value\_as\_concept\_id

This is the main table for values calculated from previously migrated values. For example if BMI is not directly available, it can be calculate from height and weight. Both height and weight would come in from a study and be stored under their respective concept’s. The calculation would retrieve those values from the CDM, calculate BMI and store it under the give concept.

* + Study\_id
  + Vocabulary\_name
  + Concept\_name
  + Function\_name: a name that maps to a CategorizationRule in categorization.py
  + To\_column
  + To\_table
* Ohdsi\_calculation\_argument
* vocabulary |  concept  |       function\_name       | arg |      argument\_name      | to\_vocab | to\_concept
* ------------+-----------+---------------------------+-----+-------------------------+----------+------------
* SNOMED     | 63467002  | best\_logical\_concept\_or   |   1 | Left BBB                | SNOMED   | 6374002
* SNOMED     | 59118001  | best\_logical\_concept\_or   |   2 | Right BBB               | SNOMED   | 6374002
* SNOMED     | 11851006  | best\_logical\_concept\_or   |   1 | mitral valvular disease | SNOMED   | 368009
* SNOMED     | 8722008   | best\_logical\_concept\_or   |   1 | atrial valvular disease | SNOMED   | 368009
* SNOMED     | 424144002 | best\_creatinine\_clearance |   1 | age                     | SNOMED   | 442407001

Tells where a calculation function gets its argument values from.

* + Study\_id
  + Vocabulary\_name: source value vocabulary
  + Concept\_name: source value concept
  + function\_name: (redundant?)
  + argument\_order: (arguably, I could use named arguments in python. TODO)
  + argument\_name: for documenting the purpose
  + value\_field: one of value\_as\_number, value\_as\_string, value\_as\_concept\_id. The others will be null. (Q: is this used on both sides?)
  + to\_concept\_name: part of FK to ohdsi\_calculation\_function
  + to\_vocabulary\_name: part of FK to ohdsi\_calculation\_function
  + from\_table: source value table
* Categorization\_function\_metadata
* function\_name    |    to\_column     | rule\_id | from\_vocab | from\_concept | from\_table  | short\_name
* --------------------+------------------+---------+------------+--------------+-------------+------------
* map\_concept\_id     | Gender           | 0       | SNOMED     | 263495000    | observation | Sex
* map\_concept\_id     | Tobacco use      | 0       | LOINC      | 39240-7      | observation | Smoke
* map\_concept\_id     | Prior PCI        | 0       | SNOMED     | 41339005     | observation | PCI
* map\_concept\_id     | Valvular disease | 0       | SNOMED     | 368009       | observation | Valv
* ranges\_to\_rank\_asc | MLHF             | 0       | LOINC      | 71938-5      | observation | MLHF

This is the main table for the extraction. It makes more sense when you think of just copying a value out as making use of the ‘identity’ function. This extraction produces an analysis matrix for use with various machine learning algorithms. The point of the functions is to allow for categorization or discretization or conversions on the way out. No concept is required for the resulting value.  
  
Example of 2-rules on the same output extracted\_name, with qualification: Anemia has different constants in the calculation for men and women.

* + Extracted\_name: a longer name for the column or phenotype in the output
  + Rule\_id: used to distinguish rules when more than one may qualify.
  + Function\_name
  + From\_vocabulary\_name
  + Comment
  + From\_concept\_name
  + From\_table
  + Short\_name
* Categorization\_function\_parameters
* function\_name    |        to\_column        | rule\_id | value\_limit | rank | from\_string | from\_concept
* --------------------+-------------------------+---------+-------------+------+-------------+--------------
* ranges\_to\_rank\_asc | Age                     | 0       |          50 |    1 |             |
* ranges\_to\_rank\_asc | Age                     | 0       |          60 |    2 |             |
* ranges\_to\_rank\_asc | Age                     | 0       |          70 |    3 |             |
* ranges\_to\_rank\_asc | Systolic blood pressure | 0       |         120 |    1 |             |
* ranges\_to\_rank\_asc | Systolic blood pressure | 0       |         140 |    2 |             |

These parameters allow different phenotypes to have different cut-off values for their different categories.

* + Extracted\_name
  + Rule\_id
  + Rank: the category number that goes with values in this range
  + Function\_name: a name that maps to the python class for each function. See CategorizationRule.\_functionFactory()
  + Value\_limit – the limit value for a particulary range
  + From\_string – the string to match. Used in functions other than the range ranking ones.
  + From\_concept\_name: the concept to match when mapping concepts.
* Categorization\_function\_qualifiers  
    
  In some (seemingly rare) cases, you want to run one of two or more functions depending on other values. Creatinine clearance for example, has a different formula to be used depending on whether the subject is male or female. This looks to see that for the subject in question, a value under the vocabulary and concept listed matches the value in one of the three value fields. The specific choice of which value field is specified in the python source for that function.
  + Function\_name
  + Extracted\_name
  + Rule\_id
  + Vocabulary\_name
  + Concept\_name
  + Value\_as\_string
  + Value\_as\_number
  + Value\_as\_concept
* Categorization\_function\_table
* function\_name |     to\_column      |      from\_table      |     from\_column      | from\_vocab | from\_concept
* ---------------+--------------------+----------------------+----------------------+------------+--------------
* x             | Withdrawal         | visit\_occurrence     | visit\_concept\_id     |            |
* x             | HF Hospitalization | procedure\_occurrence | procedure\_concept\_id | SNOMED     |
* x             | Death Transplant   | Death                | x                    | y          | z
* is\_not\_null   | Transplant         | procedure\_occurrence | procedure\_concept\_id | SNOMED     | 32413006
* is\_not\_null   | Hospitalization    | procedure\_occurrence | procedure\_concept\_id | SNOMED     | 32485007

This table lists extractions from the wide tables: visit\_occurence, procedure\_occurrence and Death

* + Function\_name
  + To\_column /extraction\_name
  + From\_table
  + From\_column
  + From\_vocabulary\_name
  + From\_concept\_name
* Events\_mapping
* from\_table   | from\_column |       to\_table       |      to\_column       | value\_vocab | value\_concept |        addl\_column         | addl\_value |                 from\_date\_column
* ----------------+-------------+----------------------+----------------------+-------------+---------------+----------------------------+------------+--------------------------------------------------
* best.xp        |             | procedure\_occurrence | procedure\_concept\_id | SNOMED      | 32413006      | prodcedure\_type\_concept\_id | 44786630   | xpdat
* best.hv        |             | procedure\_occurrence | procedure\_concept\_id | SNOMED      | 32485007      | prodcedure\_type\_concept\_id | 44786630   | hvdat
* hfaction.death | deacard     | Death                | cause\_concept\_id     | SNOMED      | 26636000      | death\_type\_concept\_id      | 38003570   | date('2015-01-01') + dthdays \* interval '1 day'
* best.adju      | adcause     | Death                | cause\_concept\_id     | UCD.Kao     | UCD-Kao-2     | death\_type\_concept\_id      | 38003570   | addate
* best.adju      | adcause     | Death                | cause\_concept\_id     | UCD.Kao     | UCD-Kao-4     | death\_type\_concept\_id      | 38003570   | addate

This is for mapping from a study’s table to person, visit\_occurrence and procedure\_occurrence.

* + Study\_id
  + From\_table
  + From\_column
  + To\_column / extraction\_name
  + Value\_vocabulary\_name
  + Value\_concept\_name
  + From\_date\_column
  + Where\_clause
  + Comment