**Always leave fields null when there is no data present**

**Location**

Q: What do we do with sources that only has region information in it?

A: Only have the Location ID and the source value. Do not force any fields into the model if we do not have it. For example, P+, we already have the City, State and Zip, but wanted to add region, however we didn’t add.

Q: What do you put in the source value when you have more than 1 field available in the source data?

A: Concatenate them into the source value

**Care Site**

Q: What is considered a care site?

A: The actually physically facility where the service will take place.

**Provider**

Q: What if a provider has multiple specialties? Do we create multiple records for the same provider or is there some logic used to determine which specialty to use if the primary specialty is not indicated

Q: What if a provider has multiple care sites? Do not load, create multiple record per provider/care site combo, pick one, or use some logic. One logic is to count the provider’s transactions and use the care site with the highest frequency

**Person**

Q: What if we have multiple gender or race

A: Do not track history, always store the most current version of the data

Q: Do we fill in the provider ID if we do not know which provider is the primary care provider

A: No, leave it blank

Q: Do we fill in the care site ID if we do not know which care site is linked to the primary care provider?

A: No, leave it blank

Q: US Data: What if we have patients over 85 or under 24 month? HIPAA requires that we cannot identify the exact age of these individuals

A: Default their year of birth to 85 or 24 months

S: If the person’s year of birth is greater than dataset’s end date, then the patient is removed from the CDM

S: Always populate the xx\_source\_concept\_id = 0 if the xx\_source\_value is null. This is needed to populate reports like Achilles

S: Do not bring in person without year of birth. Do not default to a fake date

Q: If someone has no transactions in event tables, should they still be in the Person table or should we exclude them?

**Death**

S: Privacy requirements – not allowed to show the date or reason for death. Death Date set to 1999 and reason set to 0

Q: What do you do when you have multiple death records for the same person?

A: Take the last death date

Q: What do you do when you have multiple causes for the same person?

A: If multiple death records, then take the latest record and use that cause of death. If the latest record has multiple cause of death, then create multiple records to store the multiple causes of death

Q: What do you do when you have transactions for a person after they have died?

A: Rule dependent on dataset. One dataset nullifies the death record if the person has transaction 60 days after the death date. Another dataset remove all transactional records after 90 days of death

**Event tables:**

**S:** Always following the target concept domain and insert it into the right table. Example, if the target concept has a domain of Observation, it should go into the Observation table even if the original source was a diagnosis/condition.

S: Every source value should have a source concept Id. They are created and using the 2 billion series

S: There should be no records that exist outside of the observation period start and end date. If needed, then the records should be record as observations using “history of”.

**Condition Occurrence:**

S: Condition\_end\_date should not be populated unless you know the actual condition end date. Most claims data has a service start/end date. The service end date should not be used to populate the condition end date.

S: Condition\_type\_concept\_Id – For claims, populated with the position of the diagnosis. For EMR, populate with 38000245 – EHR problem list entry

Q: Should the Visit occurrence ID should always be populated

A: In our world, it should always be populated

S: Privacy requirement: If the condition is considered a privacy issues, the record is not populated in the CDM. Example are sexual abuse or car accidents

S: Claims – When there are multiple provider IDs and a way to identify the different kind of provider, the Rendering provider is populated in this table

S: Source value – should not be storing the source value concatenated with the source description. If the source value does not already exist as a concept, create a new concept (2 billion series) and store the source description as the concept name. Then refer to the new concept in the source\_concept\_Id field. Also, if the source value is already in the concept table, we standardize the format of the source value. Example, ICD9/10 codes – some sources do not store the dots and some source do. We store them with dots.

S: Always remove duplicate condition records (especially with claims based data). The only positions that are recorded more than one is the admitting and primary diagnosis. All others, use the first position

**Procedures:**

S: In claims data there may be a field procmod (procedure qualifier - additional information about procedure) that is related to pproc (primary procedure). Usually proc goes to ~5 cdm domains (Condition, Procedure, Measurement, Observation, Device), but only in PROCEDURE\_OCCURRENCE and OBSERVATION it is possible to save qualifier (procmod) in the same record with proc. In such case create additional records in event tables for all procmod codes according to their domain\_id, then create records in FACT\_RELATIONSHIP to link cdm records created for procmod and proc from the same source records (with relationship\_id = 0).

Q: When do we de-dup procedures?

Q: What is the correct type\_concept\_id?

Q: Do we populate quantity? When we do not have quantity what is the default? NULL or 1?

**Observation**

Q: When do you use value as a string, value as a concept id and value as a number?

Q: what is the correct observation type concept to use for claims data versus EHR data?

Q: How is qualifier populated?

S: Needs to be calculated based on data set’s eligibility rule (Observation Period)

**Measurement**

Q: When do you use value as a concept id and value as a number?

Q: How do we standardize the data?

**Drug Exposure**

S: If there is no start date, then we do not include the record

S: End date should be calculated using days supply. If there is no days supply we give it a default of 30 days.

Q: Do we de-dup drugs? Example someone can be given the same drug twice in one day if they are in the hospital. What if it comes from a Pharmacy?

Q: What if you have contradictory information between end date, sig and days supply?

S: All drugs should be tied to a visit

**Device Exposure**

Q: What are the rules for de-duping?

Q: what do you populate when no quantity information? NULL or 1?

Q: Should visit occurrence always be populated for Device Exposure?

**Specimen**

**Payer Plan Period**

**Cost**

S: We do not convert currencies