In large part, the following describes what happens in /proj/DaltonLab/projects/p0013/progs/04\_Harmonizing\_CCF\_and\_MH\_data.Rmd

Essentially, this document creates views that combine the analogous CCF and MH data sets into single tables (e.g., combines MH’s and CCF’s meds tables into a single table).

Code chunk “deaths”.

* Creates DL\_NEOCARE.NEOCARE\_COHORT\_DEATHS\_V on Teradata.
* The entirety of DL\_NEOCARE.CCF\_COHORT\_DEATHS\_V is concatenated (UNION ALL) with what is essentially a subset of DL\_NEOCARE.MH\_COHORT\_DEATHS.
  + DL\_NEOCARE.CCF\_COHORT\_DEATHS\_V is created in 02\_CCF\_cohort\_views.Rmd (see CCF\_cohort\_creation.docx).
  + DL\_NEOCARE.MH\_COHORT\_DEATHS
    - Created in \_\_\_ (see MH\_cohort\_creation.docx).
    - Filtered to include only study\_ids whose record in DL\_NEOCARE.STUDY\_ID\_KEY has from\_ccf = 0. As a result:
      * CCF\_COHORT\_DEATHS\_V is the sole source of death data on NEOCARE patients who had encounters at CCF.
      * MH\_COHORT\_DEATHS only informs death data of MH-only patients.
    - Filtered to exclude records wherein the death date occurs before the first encounter in MH\_COHORT\_ENCS.
* Columns:
  + study\_id
  + death\_date
    - From CCF\_COHORT\_DEATHS\_V: death\_date\_calc renamed.
    - Taken as-is from MH\_COHORT\_DEATHS
  + UnderlyingCOD\_ICD10
    - Taken as-is from CCF\_COHORT\_DEATHS\_V
    - From MH\_COHORT\_DEATHS: column “cause” renamed.
  + UnderlyingCauseOfDeath
    - Value
      * “CV Death” if UnderlyingCOD\_ICD10 begins with I00 through I79
      * Missing if UnderlyingCOD\_ICD10 is missing.
      * Otherwise, “Non CV Death”
    - Taken as-is from CCF\_COHORT\_DEATHS\_V.
    - From MH\_COHORT\_DEATHS: calculated from column “cause”

Code chunk “procs”

* Creates the view DL\_NEOCARE.NEOCARE\_COHORT\_PROCS\_V
* Columns: study\_id, ConceptID (which is UMLS ConceptID), and proc\_date.
* Concatenation (UNION ALL) of what is essentially all CCF procedures with what is essentially all MH procedures
  + CCF procedures
    - DL\_NEOCARE.STUDY\_ID\_KEY inner joined to CCF\_QHS\_PROCEDURES on CCF patient ID, so that only CCF procedures performed on NEOCARE cohort patients remain
    - ConceptIDs were already present in this table as-is.
  + MH procedures
    - DL\_NEOCARE.MH\_COHORT\_PROCS had CPT and CDT codes instead of UMLS ConceptIDs, so these had to be converted. This was done by inner joining the table with a key table relating each possible CPT/CDT to its preferred UMLS ConceptID. This table was a subset of QHS\_UMLS\_V.Atoms, created as follows:
      * The table was first filtered to only include rows whose VocabularyAbbreviation was ‘CPT’ or ‘CDT’ and whose TermType did not begin with ‘ET’.
      * Remaining rows were partitioned by Code (which could be either CPT or CDT). Each partition was sorted firstly so that those with VocabularyAbbreviation = ‘CPT’ came first, then so that those with IsPreferred = ‘Y’ came first, then so that those with TermType = ‘PT’ came first. Then, only the top row for each partition was kept.