This file describes most of the assembling of the raw MH data as well as its processing for transfer to CCF. It describes the actions of two files in the folder on the MH server named /CHRP/NEOCARE/datawork/mh\_data\_prep\_for\_transfer

* 01\_process\_mh\_sas\_files.Rmd
  + Reads in the raw SAS files and merges each one with the study ID key (viz. /NEOCARE/Corrected Data-MH Jan 2019/mh\_study\_id\_key.rds, created at the end of **deduplication – see documentation**).
    - The SAS files are located in /NEOCARE/Corrected Data-MH Jan 2019/SAS\_neocare\_files
    - Merging was based on the MRN and patient ID columns
  + Concerning the demographics, encounters, ICD9 diagnoses, ICD10 diagnoses, and medications:
    - Contains a lot of code that calculated patient ages per row. This code is now obsolete because an IRB change allowed us to send dates of events.
    - Otherwise, these were essentially converted from .sas7bdat to .rds and written into the folder /NEOCARE/Corrected Data-MH Jan 2019/mh\_cohort\_age\_calculated
  + The section processing the labs data redone later since LOINC needed to be added (a key table was left-joined in). The results were written as /NEOCARE/Corrected Data-MH Jan 2019/to\_ccf/2019-11-14/mh\_labs\_fixed\_loinc.rds. As a result, code chunks in As a result, code chunks in 02\_write\_mh\_data\_to\_transfer\_to\_ccf.Rmd that process and write labs data are now obsolete.
  + The section processing the procedures data was also redone later since it needed to be re-pulled from Epic. The results were written as procs1.rds and procs2.rds into the folder /NEOCARE/Corrected Data-MH Jan 2019/to\_ccf/2020-01-15. As a result, code chunks in As a result, code chunks in 02\_write\_mh\_data\_to\_transfer\_to\_ccf.Rmd that process and write labs data are now obsolete.
* 02\_write\_mh\_data\_to\_transfer\_to\_ccf.Rmd
  + The demographics, encounters, ICD9 diagnoses, ICD10 diagnoses, and medications data were further processed in this file. Only necessary columns were extracted, and the data were written as .rds files into the folder CHRP/NEOCARE/Corrected Data-MH Jan 2019/mh\_cohort\_age\_calculated/to\_ccf/2019-06-17
  + Concerning only the smoking and GEOID data, this file also reads in the raw SAS files and merges each with the study ID key (see above). Some light processing is performed, and then these data are written as .rds files into the folder CHRP/NEOCARE/Corrected Data-MH Jan 2019/mh\_cohort\_age\_calculated/to\_ccf/2019-06-17
  + Code chunks processing procedures and labs data are now obsolete (see above).

Both of the above files’ resulting .rds data sets were transferred from the to\_ccf folder to the Cleveland Clinic via FileZilla. In addition, MH financial class data was transferred to CCF as a raw .csv file with light processing.