**Dia\_Comorbidities\_Over\_Time.ipynb**

Input:

* The 4 diagnoses files:
  + Raw Data/First/txt/KS185\_20200918\_114153\_Dia.txt
  + Raw Data/Second/txt/KS185\_20200918\_114153\_Dia.txt
  + Raw Data/Third/txt/KS185\_20200918\_114153\_Dia.txt
  + Raw Data/Fourth/txt/KS185\_20200918\_114153\_Dia.txt
* Patients we want comorbidities for
  + Processed data/singlerp.csv
  + Processed data/df\_pathology\_biopsy\_final.csv
  + Code/Diagnoses/empis\_need\_comorbidity.csv
* Mapping from LMR codes to ICD9 codes
  + Raw Data/lmr\_icd9map.csv
* Intermediary Input from the output of our R file (see below):
  + Code/Diagnoses/charlson\_singlerp.csv

Output:

* Intermediary output as input of our R file (see below):
  + Code/Diagnoses/diamerged\_singlerp.csv
* Final output containing patients, their Charlsen comorbidities, score and weighted scores “rolling” across time
  + Processed Data/dia\_comorbDL.csv

How it works:

* The diagnoses files contain diagnoses in terms of ICD9, ICD10, LMR, among other Codes for patients on certain visit days. I subset our diagnoses file to our patients of interest (in singlerp, biopsy\_final, and empis\_need\_comorbidity). I also map certain LMR codes to ICD9 codes using lmr\_icd9map.csv. Then, I export the file with diagnoses codes and our patients of interest to R, where we can use a R package to compute Charlsen comorbidity scores and weighted scores from ICD9 and ICD10 codes.
* After obtaining the Charlsen comorbidity scores from the R script for each observation (patient-day), I combine the ICD9 and ICD10 comorbidity scores under each patient-day, aggregating them so there is one score for each patient-day.
* Then, I compute a “rolling” comorbidity score and weighted score. I sort the data by patient and date, then aggregate comorbidities as always occurring in the future after they have been seen once in a patient (this is done by iterating through all the rows of the dataset, then updating a score as the sum of the previous visit’s cumulative score and the current visit’s score). Then, each patient has weakly increasing aggregate comorbidity scores over time.

**dia\_comorbidity\_singlerp.R**

Input:

* The data obtained from the previous Jupyter notebook file, with patient-date observations we want to compute comorbidities for
  + diamerged\_singlerp.csv

Output:

* Comorbidity indicators, scores, and weighted scores for each patient-date observation
  + charlson\_singlerp.csv

How it works:

* We use the R package “comorbidity” to calculate comorbidity indicators, scores, and weighted scores from ICD9 and ICD10 codes. We import the data, subset the data into those with ICD9 codes and those with ICD10 codes, then compute comorbidities separately for each subset. Then, we output the appended subsets of data back into the Jupyter notebook (see above) for further analysis.