**Report for Coding Exercise**

**Summary**

For this exercise, I ran a logistic regression on the dataset to determine the likelihood that a patient admitted for CABG surgery would be readmitted within 30 days. I trained the model on the full dataset and then ran the trained model on the subset of patients admitted for CABG surgery. I found that it is extremely unlikely that a patient would be readmitted within 30 days with a likelihood of 0.272. In fact, none of the 1361 patients admitted for CABG surgery in this dataset were readmitted within 30 days.

I chose to use a logistic regression because the dependent variable (whether a patient was readmitted) was binary, which made logistic regression an appealing choice. This model also works best with large sample sizes. This model assumes that each of the entries are independent with no collinearity among the independent variables (patient identity, age, systolic blood pressure, and the number of comorbidities). For this dataset of randomly generated numbers, a few simple scatterplots were enough to show a lack of collinearity between these variables.

**Some oddities in the data**

For this exercise, I was provided with a csv file of randomly generated data. This resulted in a few amusing oddities in the dataset.

1. Every time patients were admitted to the hospital, their age changed, with individuals getting younger and older at equal rates with no correlation with the date of the admission.
2. One patient (ID 2777) was admitted 10 times and given 49 unique diagnoses!
3. The ICD-9 codes were all four digits, which excluded many possible diagnoses, including known comorbidities, from the dataset. Though this may have been intentional, for ease of data generation and code review.