**Problem Statement**

For this data challenge, you are given a real-world dataset from electronic health records of patient diagnosed with a specific condition and prescribed either Drug A or Drug B for the treatment. The patients are then monitored for the occurrence of a specific event after the start of the treatment.

The following CSV files are provided:

* Patient\_characteristics.csv – contains the information about socio-demographics of patients, diagnosis, lab values, and other existing therapies patients are on
* Event\_duration.csv – Contains information about if the event happened to a patient or if patient was censored and the durations for each case from the start of therapy

The goal of the specific task is to compare the real-world efficacy of the two drugs by comparing the risk of events using a method like survival analysis. Do keep in mind that as this data comes from EHR and not a clinical trial, the two groups may not have balanced patient characteristics which may confound the results of your analysis. You may also have to think about how you would measure and reduce the impact of this confounding in your analysis approach.