

ORIE4741 Proposal

Hao Qiu(hq54), Yuting Zhao(yz2632), Yuze Wang(yw2272)

September 25, 2019

1 Introduction and Significance

Different hospitals specialize in different medical fields. Local and central hospitals may also focus on startlingly distinctive medical conditions in terms of types of diseases, level of severity, and etc. Therefore, people may choose certain hospital over others according to their needs. In most cases, patients who have the most urgent conditions tend to choose the best hospitals, either locally or even regardless of the distance. In general, we would answer several questions including:

1. Given a medical condition, which hospital should a patient visit?
2. Is the length of stay correlated with the level of severity?
3. Does certain medical facilities receive predominately number of patients with similar disease?

We plan to choose analytical approach to study patients' preferences in choosing hospitals based on medical condition, severity of illness, length of stay, total costs and etc. Our goal is to make the hospital selection easier for potential patients using machine learning algorithms. We could redirect the patients with certain medical conditions to the optimal choice of hospital so that the hospitals may also achieve the highest efficiency of time and cost.

2 Data Sets

We will be mostly using data provided by Statewide Planning and Research Cooperative System (SPARCS) and published on health.data.ny.gov. SPARCS records treatment and cost details for a large number of hospitals in NY State. This data set contains 2.54 million entries and 34 different columns. Some example of columns are Hospital County, Gender, Race, and Risk of Mortality. We may also incorporate alternative data sources from The Medical Expenditure Panel Survey (MEPS) or Practice Fusion.

- <https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/u4ud-w55t>
- <https://meps.ahrq.gov/mepsweb/>
- <https://www.kaggle.com/c/pf2012>