Problem Statement & Significance

If you need medical help for a specific condition, where do you go? If you have a health insurance plan that dictates a set of in-network providers, which one should you choose? There is a wealth of public patient discharge data in New York State, and we would like to identify which hospitals are best to go to given a patient's specific injuries or ailments. It poses a challenge to define a goal, as there are different indicators for a "success" case, from survival rates, service rates to readmission rates. Apart from data analysis, some qualitative research is required to clarify our goals. In general, below are some questions we would like to answer:

1. Is it better to go to a large hospital with a large volume of discharges, or a smaller hospital with fewer discharges and a smaller staff?

2. Do more expensive hospitals offer better care?

3. Is cost of procedure correlated with quality?

4. To what extent does a hospital charge correlate with the cost to provide that care?

Additionally, we would like to see how quality of care differs among counties with varying income levels. In an emergency, a patient most likely will be sent to their nearest trauma center, but will they receive a sub-standard care if they live in a county with a high poverty level?

On a secondary level, from a hospital administration point of view, understanding which procedures or diseases the hospital has a good success rate with, or otherwise, is important for formulating future strategies. Moreover, answering these questions allow us to evaluate the effectiveness of, for example, the value-based model versus the volume-based models.

All in all, through answering the question - "which hospital to go", we aim at understanding current health care model from social perspectives and adminstration perspectives, in hope to suggest some favorable strategies that the hospitals, insurers or even policy makers can consider.

Data Sets

We will be mostly using data published on health.data.ny.gov that was collected by SPARCS(Statewide Planning and Research Cooperative System) as a start. SPARCS operates under the Department of Health, and aggregates data from almost all medical care facilities in NY state. They cross check submitted data with patient records and other registries to ensure that records aren't falsified or contain egregious errors. To see if there are connections between quality of care and geographic area and income levels, we will join that data with income level data and poverty levels by county as published by the Census.

- https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/u4ud-w55t

- http://www.ers.usda.gov/data-products/county-level-data-sets/poverty.aspx

- https://www.census.gov/did/www/saipe/data/interactive/saipe.html?s_appName=saipe&map_yearSelector=2014&map_geoSelector=aa_c