BEM Check-In 2

Advith Chelikani, Nick Haliday, Logan Cross

Below are several hypotheses we wish to test with our data. They are largely concerned with overutilization (receiving more medical treatment that necessary).

**Overall concern**: Since money-related healthcare data is often propriety and kept under lock and key, we are having trouble finding relevant data that we could use in our project.

1) What regions have excessive costs related to the over prescription of expensive tests, surgeries, and other high-powered medical treatments (also known as overutilization)?

**Expectation**: We expect that large cities and suburbs of large cities, which will likely have higher quality hospitals, have higher incidence rates of overutilization.

**Comments**: We may limit this to regions within a specific state or set of states due to the sheer size of the data set.

**Current Status**: Currently, we are testing our procedure on the state of Delaware (because it has only ~4,000 entries vs. the 180,000 entries for California)

**Procedure**: Firstly we noticed that the Factual data does not contain information on county (it has locality and neighborhood, neither of which are county). We require county data in order to be able to join the Factual data set with other data. Using a reverse geocoding API, we have modified the Factual data (for Delaware for now) to be anonymized and include county (which we derived from the latitude, longitude value). Then using open source data on costs of medicare by county and mortality rates by county we are building a classifier that determines whether a person with a given location is experiencing over utilization. Something we still need to work on is determining a bound for overutilization by county. Also, there are many other factors that likely contribute to overutilization that we haven’t taken into account. Among these factors are experience level of doctors and type of doctor (dentist, pediatrician, etc.). With respect to experience, unfortunately the Factual data is missing this information for >50% of all records, so we will have to find a different source.

**Preliminary Findings:** Our tests with Delaware were not very conclusive since there are only three counties in Delaware. We plan to repeat this process with larger more prominent states and hopefully will see better results.

2) Does overutilization scale with the experience/proficiency of doctors in a locality? If so, to what extent?

**Expectation**: We expect that overutilization does scale with experience and proficiency. The more qualified a doctor is, the more likely they receive wealthier patients who would be more willing to spend money on healthcare (i.e. money is not a major concern for them with regards to health).

3) Does overutilization scale with salary of doctors?

**Expectation**: We expect that overutilization does scale with salary of doctors.

4) Use linear regression to identify explanatory variables accounting for excessive use of low-value services. In the recent Choosing Wisely Initiative (http://www.choosingwisely.org/doctor-patient-lists/), dozens of low-value services were identified, as defined by an expensive service that does not improve or worsens health outcomes. An example of a low-value service is the prescription of opioids for minor migraines (the pills are expensive, do not treat the underlying problem, and put a patient at risk for addiction). For this regression analysis, demographic factors within a county, specialist and primary care physician densities will be input as variables. As many demographic factors are correlated and/or unrelated to health, ridge or LASSO regression will be used.

**Expectation:** The amount of low-value services will be correlated with the percentage of minorities in a region and a higher specialist to primary care ratio. Since primary care physicians are more likely to promote preventative measures (ie. exercise more and eat healtier) rather than prescribe treatments and surgeries that they are trained for, we hypothesize that the specialist to primary care ratio within a region will be highly correlated to excessive use of low-value services. Additionally, doctors in lower-income areas with minority demographics are less likely to be familiar with innovative cost-effective methods that are used in place of low-value services.