BIO 260 / CSCI E107 Final Project Proposal

This is the form your team must fill out for the Final Project Proposal (due April 8, 2016 at 11:59pm EST). Based on your proposals you will be assigned a TA to your team who will guide you through the rest of the project. You will schedule a project review meeting with your TA. Make sure all of your team members are present at the meeting. Online students can schedule a Skype meeting or Google Hangout with their TA.

\* Required

Title of your project proposal \*

Ambulatory Care : Who uses it and Why ?

Team Member 1's Name \*

Vidya Bhalodia

Background and Motivation \*

Ambulatory and urgent care facilities are starting to become common alternatives to emergency rooms and hospitals for many ailments and procedures. During my 10 years working in the operating room, I’ve seen procedures that required a 3-4 day inpatient stay now being performed on a completely outpatient basis at surgicenters. Down my street, there are advertisements informing potential patients that if they ever have a minor emergency, they don’t necessarily have to wait hours in the ER, when they can just stop by the local urgent care center.

I expect that ambulatory care will become an increasing and important part of the USA’s healthcare future. With that in mind, I am curious to understand how these facilities are currently being used. What are the characteristics of patients who use these facilities ? What types of services are they typically receiving ?

By understanding who utilizes these ambulatory care facilities, and how, we can get an insight as to how best allocate our limited healthcare dollars.

Project Objectives \*

What are the scientific and inferential goals for this project? What would you like to learn and accomplish? List the benefits. What are some optional features (features or calculations which you consider would be nice to have, but not critical)?

Descriptive Analytics : I expect to do basic exploratory data analysis and simple visualizations (histograms) on the following parameters (individually as well as by region and/or hospital ownership)

* Who is using these facilities : By gender, age, insurance providers
* Why are they using these facilities :
  + What are the most common symptoms / diagnoses ?
  + How does the healthcare provider rank the urgency of the symptoms ?
* What is done during these visits :
  + How many tests / procedures are performed ?
  + Medications Provided ?
  + Which healthcare providers do they see ?
* How long are the visits : Wait time, duration
* How many patients left (or died) before receiving treatment ?

Time Series / Trends :

The data is provided over multiple years. I would like to look at several trends over time:

* How has wait time and visit duration changed over the years ?
* How has the number of tests performed changed ?
* How has the number of medications provided changed ?

Hypothesis Testing and Linear Regression :

* Is there a relationship between Pain Scores and Wait Times
* Is there a relationship between Reported Urgency and Wait Times
* Is # Procedures Performed dependent on Insurance Type
* Do patients with chronic conditions (diabetes, chf, HIV, etc) have more visits to the ER than patients who don’t ?

Model Building :

I would like to create a model that predicts whether a patient will be admitted to a hospital or other inpatient facility based on the results of their ER visit (other than diagnosis codes). I would like to use the parameters of visit duration, # diagnostic tests, # of procedures performed, # of medications ordered, age, gender, previous ER visits, previous hospital discharge,

What Data? \*

The data is available online : <http://www.nber.org/data/national-hospital-ambulatory-medical-care-survey.html>

The data is available for the years 1992 - 2011, which should give me enough data to create a time series.

Design Overview \*

List the statistical and computational methods you plan to use.

* Descriptive Statistics : mean, median, standard deviation, standard error, outliers.
* Histograms :
* Time Series Plots
* Linear Regression
* Possibly Logistic Regression
* Predictive Modeling (likely GLM, but perhaps KNN).
* If I’m feeling super duper ambitious, I may try to create an ROC for admissions thresholds.

Schedule / timeline \*

4/16 – Process and Clean Data

4/23 – Exploratory Data Analysis, Descriptive Statistics

4/30 – Regressions and Model Building

5/4 – Done