Preventing Hospital Readmission After Diabetic Encounters

Michael Gat

Capstone Project Proposal

for

General Assembly Santa Monica
Data Science Immersive Program
Summer 2016

Objective

Create a model to identify individuals with highest risk of unscheduled hospital readmission after a diabetic encounter

- Ensure those at highest risk are identified and offered focused care
- Identify the circumstances that are most strongly associated with readmission so those can be specifically addressed when present
- The key is in the patient history, which when properly encoded offers an excellent opportunity to apply machine learning techniques.

Some background information

- Diabetes
  - Direct cause of ~9% of U.S. healthcare costs
  - Affects ~10% of the population
  - 7th leading direct cause of death
- Hospital Readmissions (patient has to come back)
  - < 30 day readmission rate is a key measure of quality of care
  - A big driver of costs, over $20b for Medicare program alone
  - There are significant penalties for high readmission rates
  - ***Often preventable!***

Why this study?

- Diabetes-associated readmission data offer a real opportunity to improve care and reduce costs
- The impact of condition-specific treatments on readmission statistics has not been studied in depth.
  - Only one paper exists relating to diabetes, and it focused on the impact of only one clinical measure (HbA1c)
- A working model is likely to be valuable
  - Especially if it can be applied to other medical conditions

Audience/stakeholders

- CMS
- Hospital administrators
- Clinical staff
- Insurance companies
- Researchers (academic and commercial)
- Patients and patient advocates

Primary Data

UCI Machine Learning dataset

- 100,000 treatment records with <=30 day, >30 day and no readmission
- Categorization dataset with definitions of key variables
- Data collected 1999-2008
- Diagnosis encoded with ICD-9
- Include anonymized information about the patient, treatment, attending physicians, length of stay, test results, drugs administered and more
- https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008

Secondary Data

- ICD-9 codes
- National and regional readmission statistics for comparison
- Readmission by ICD-9 coding for comparison
- Anonymized patient data for other conditions
  - If I complete a useful model with the public dataset, I may be able to obtain an academic license to additional data.
  - Would be great to get a breakdown by geography/facility

Further references

- *Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records*
  http://www.hindawi.com/journals/bmri/2014/781670/