



*GA – Data Science Capstone Proposal*  
*Predicting Patient Readmission for Chronic Illnesses*

*Ben Doan*



## Overview

---

Chronic illness readmission is one of the leading cost drivers for healthcare providers in the United States, for several reasons:

- Complexity of care required to treat ongoing conditions
- Resource costs for specialists required to treat chronic conditions
- Patient transport and facility transition costs during readmission
- Reimbursement penalties associated with patient readmission within 30 days.

In my current client engagement at a major hospital system in Central Pennsylvania, I am working as the PM for a “High Risk Triage Team,” that works to identify patient cases that are at high risk for readmission\* after being discharged.

Once patients are identified, they are tagged and then assigned to a specific team of clinicians (mainly doctors and nurses) that have specialized skills to properly treat and discharge.

*\*In this case, “high risk” refers to the likelihood that the patient will be readmitted to the hospital within 30 days.*

## Problem Statement

---

The current case identification process is done at a macro-level, using disease/condition type as the major identifying factor for “High Risk” patients.

As a result, many patients with minor presentations of symptoms or less advanced progressions are identified as “High Risk”, even though they are not. Thus, “low risk” patients are assigned highly skilled hospital resources, when the treatment can be administered by more generalized practitioners.

Additionally, the tail-end of the identification process is done manually in 4-hour screening meetings, twice per week.

## Hypothesis

---

Machine learning algorithm can be used to classify patients into “High Risk for Readmission” or “Non-High Risk for Readmission” categories using data on care facility, patient demographic, and lab/test results for a specific chronic illness. \*

*\*Chronic illness will be determined by client clinical care manager, after patient data is anonymized and released for the purpose of this project*

## Assumptions

---

- Data will be complete and comprehensive
- Whether or not the patient was readmitted within 30 days will be noted in the data, or can be computed using discharge date and next admission date



- Data is chosen from a sample of patients within the hospital system that all present with the same condition
- Data can will be presented in .csv format
- Results will be representative of my client's patient population only, and **not** the larger American patient population at large
- Results will only be representative of the selected chronic illness only, and **not** chronic illnesses at large

## Risks

---

- Data will have "Comment" field that may/may not contain important information that will have to be parsed out
- Data may contain sensitive patient information that cannot be shared outside purposes of this study
- Data may have missing data from delayed tests, procedures, or lab results
- Some of the data (patient race, age, etc.) may be skewed due to the patient demographics

## Data Set

---

- Data will be provided by client's clinical care manager
- Data will include anonymized information and health data for patients with a specific chronic condition
- Fields will include patient demographic information, facility information, lab/test results, previous discharge information, and admission/readmission information
- Chronic condition will be selected by client's clinical care manager by end of July 2018.