ADS Fall 2017 – Project Proposal

# Team Info

Team name: Doctors without Boundaries

Team Members:

* Rafael Buitrago
* Chloe Lee
* Stephanie Bruno

# Need for the Project

During the last several years, the number of overdose and opioid-related deaths in the United States has increased dramatically. This increase in overdose and opioid-related deaths, named the opioid epidemic, has created a great deal of interest in identifying the causes and solutions to this crisis. These questions aim to identify what factors most contribute to the opioid epidemic; how it can be mitigated; and what, if anything, can the available abundance of associated data do to help overcome this crisis.

There are multiple parties who have a stake in reducing the number of opioid-related deaths. These range from the communities affected by it; the clinicians who treat these patients; the Commercial and Medicare and Medicaid insurance payers who pay for the expenses these patients incur for treatment; and the government and tax payers who are ultimately responsible for the financial burden that the opioid epidemic is increasingly bringing upon our society.

**Our group project aims to identify the subset of healthcare practitioners (HCPs) that, based on their prescribing habits and other personal attributes, are most likely to incur into indiscriminate prescribing of opioids.**

While there are many factors contributing to the increasing rise of prescription opioids utilization, we will direct our analysis towards constructing a predictive model that may be utilized to identify the HCPs, and associated circumstances, that will likely prescribe opioids to patients with a higher frequency than in valid situations.

We believe that a predictive model that identifies the most prone-to-prescribe HCPs could be of utility to the government, insurance companies, and the CMS systems to reduce overdose and addiction treatment costs, and most importantly, to eventually reduce the number of deaths associated with the opioid epidemic in the United States.

# Hypotheses

Does a collection of HCP-prescribing patterns such as the combination of commonly prescribed drugs with certain opioid medications, or opioid-prescribing frequency, provide associative links that can determine the pool of characteristics that constitute the healthcare practitioner's profile that insurance companies and the government, including the CMS centers, may utilize to reduce the costs associated with the continuing opioid epidemic?

# Technical Approaches to be Taken

To develop our model, we will use Python, R to do initial data cleansing, Tableau for data exploration, as well as to try out different data mining techniques that we have learned in previous courses and those we expect to learn throughout this course. For visualizations, we will use either Tableau, Shiny, or Power BI.

# Data

We will use a dataset sourced from Kaggle.com, [U.S. Opiate Prescriptions](https://www.kaggle.com/apryor6/us-opiate-prescriptions). This dataset contains HCP data from CMS (Centers for Medicare & Medicaid Services) which has an identifier for each HCP along with several provider attributes such as gender, region, urban or rural setting, specialty, and years practicing for approximately 25,000 providers. In addition, the dataset lists prescription counts for each prescriber for all drugs they prescribed and which claims were payable under CMS. Lastly, there are counts of the number of prescriptions per provider of generic vs. branded drugs. In addition to this dataset, we will also use the [Medicare Part D drug categories dataset](https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/PartD2015.html) to identify the different types of drugs that are prescribed, such as opioids, antipsychotics, etc.

# Project Execution Risks and Mitigation Strategies

One of the limitations of this project is the dataset itself. Because it comes from only Medicare Part D beneficiaries, it may not be representative of the U.S. population as a whole. It's also important to note that the purpose of this project is not to lay blame for the current opioid epidemic at the feet of health-care providers (despite our project team's name), so we must be careful in how we analyze and present our findings.

# Goals

The goal of the project is to identify health care providers who are likely to be high-volume prescribers of opioids, based on the [U.S. Opiate Prescriptions](https://www.kaggle.com/apryor6/us-opiate-prescriptions) dataset, the Part D Drug Category List, as well as the real-world dataset. In doing so, this model could provide stakeholders with important information when making decisions about strategies for managing the current opioid crisis.

# Deliverables

The final deliverables for the project will be a report detailing the data used, the methodology we took, and the final model built to predict those providers likely to be high-volume prescribers of opioids. We will provide an analysis of our final model. Additionally, we will develop visualizations to better support and demonstrate our findings.