The 2016 Opioid Prescription Analysis

Opioids are a broad group of pain-relieving medications derived from opium. The common types of Opioids drugs include Codeine, Hydrocodone, Morphine, Oxycodone, Hydromorphone and Fentanyl, which can have therapeutic effects, but can also trigger opioid addiction[1]. The overdependence of opioids by consumers and the concomitant over-prescription of opioids by some providers have resulted in a colossal opioid crisis. More specifically, in 2015, 33,091 people died from opioid overdoses, of which 15,281 deaths were attributed to commonly prescribed opioids[2]. Also, according to Center for Disease Control and Prevention, the number of prescription of opioids per 100 people is 66.5, which is extremely high[3]. To address these concerns, it is meaningful to investigate how the heterogeneity in provider's state and speciality influences the proportion of providers prescribing opioids, proportion of opioid-related claims and the number of day's supply of all opioid drugs (opioid day supply).

This analysis considers two relevant Medicare Provider Utilization and Payment Data: (1) 2016 Part D Prescriber and (2) Part D Prescriber Summary Table CY2016. The two datasets were disseminated by the Centers for Medicare & Medicaid Services (CMS) for public use[4]. To begin, I selected the distinct providers by subsetting the unique National Provider Identifier (NPI) for both datasets and merged the two datasets according to the intersection of their unique NPI. The resulting dataset consists of 893,160 observations and 84 columns; each observation presents primarily the information such as provider's NPI, state, city, specialty type, total days of supply for all drugs or only opioid-related drugs, and the proportion of claims relevant to opioids.

Prior to my analysis, I removed all rows with missing data (NA) in the columns related to opioid prescription. Additionally, I removed the providers in the nine states with less than 500 providers, leaving the dataset with 52 states, including the 50 actual states, Washington D.C. and Puerto Rico. Further, I added a categorical variable "Opioid" and assigned 1 to the providers with opioids prescription (opioid days of supply $\neq 0$) and 0 to the ones without opioid prescription. After cleaning the data, the dataset reduces to 625,484 rows.

In the exploratory data analysis, I yielded some important descriptive statistics about the opioid prescription in 2016: (a) the average opioid claim per provider was 122.7; (b) the overall proportion of providers who prescribed opioid was approximately 68%; and (c) the opioid day of supply is heavily right-skewed, with the mean value slightly more than 4,000 and the median of only 751. To further examine the data in terms of providers' state and specialty, I found that among the 130 specialties with opioids prescription, providers specialized in Interventional Pain Management and Pain Management prescribed the most days of opioids (45,508 and 35,890). However, providers specialized in Pediatrics and Rheumatology prescribed the opioids with largest volume (32.05 and 30.11), which is defined by $\frac{\textit{day of opioid supply}}{\textit{total opioid claims}}$. Moreover, the figure shows that providers in Alabama prescribed the most days of opioids on average (8,065), which is quintuple to Washington D.C. (1,605), where the providers prescribed the lowest days of opioids. Furthermore, the figure shows that the highest percentage of providers prescribed opioid drugs in 2016 was 80.2% in Arizona, whereas the highest proportion of opioid-related claims witnessed in Delaware (32.9%). In short, difference in provider's state and speciality contributed tremendously to the disparity of opioids prescription represented in the data.

To further investigate the patterns of the opioid day of supply and proportion of providers prescribing opioid in 2016, I fit two separate fixed effect models. The first model examines the effects of provider's state and the log-transformed non-opioid day of supply to the log-transformed opioid day of supply, while the second model examines the effects of provider's state and the log-transformed non-opioid day of supply to the proportion of opioid-related claims. The model specifications are:

$$
\begin{aligned}
log(opioid\ day) = 0.133 + 0.617 \cdot log(nonopioid\ day) + 0.197 \cdot I(Alabama) + \ldots - 0.81 \cdot I(Puerto\ Rico) \\
+ 3.79 \cdot I(Interventional\ Pain\ Management) - 3.32 \cdot I(Respite\ Care)
\end{aligned}
$$

$$
\begin{aligned}
P(opioid\ claims) = 6.051 + 1.092 \cdot log(opioid\ day) - 0.225 \cdot I(Colorado) + \ldots - 10.159 \cdot I(Puerto\ Rico) \\
+ 47.047 \cdot I(Thoracic\ Surgery) - 11.113 \cdot I(Community\ Health\ Worker)
\end{aligned}
$$

While it is difficult to truly predict the log-transformed opioid day of supply and the proportion of opioid-related claims, these two models give an approximate estimation of the responses, based on the baseline provider state Alaska and specialty Acupuncturist. Remarkably, these models show that (a) a unit increase of log(non-opioid day supply) is associated with 0.617 unit increase of the resulting log(opioid day supply), (b) Alabama and Colorado respectively contributed the greatest in state effect for the outcomes, while Puerto Rico contributed the least, and (c) providers specialized in Interventional Pain Management and Thoracic Surgery are expected to prescribe the most days of opioid supply and the most proportion of opioid-related claims, respectively. The R^2 values for the fixed effect models are respectively 0.702 and 0.643, indicating that more than three fifth of the variability of the dataset has been captured by fitting the fixed effect models. Finally, I wanted to look closer to the random effects by fitting the random effect models. However, it is unfortunate that the data matrix for interaction terms is too huge to be computationally feasible by the statistical software R.

To capture more variability of the data, it is beneficial to incorporate the demographic data about the patients into the fixed effect and mixed effect models. In addition, employing cross-validation, shrinkage and knockoff methods in the dataset may also help increase the versatility of my analysis since they are usually more powerful tools to examine the data with both qualitative and quantitative values.

References:

1. I Want to Change My Life. (2018). *Opioids - Opiates: Addiction, Withdrawal, Crisis, Recovery Facts*. [online] Available at: https://www.addictionsandrecovery.org/opioid-opiate-recovery.htm [Accessed 11 Oct. 2018].
2. YouTube. (2018). *The agony of opioid withdrawal — and what doctors should tell patients about it | Travis Rieder*. [online] Available at: https://www.youtube.com/watch?time_continue=349&v=WhpAYw9kCt8 [Accessed 11 Oct. 2018].
3. Cdc.gov. (2018). *U.S. Opioid Prescribing Rate Maps | Drug Overdose | CDC Injury Center*. [online] Available at: https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html [Accessed 11 Oct. 2018].
4. Data.cms.gov. (2018). [online] Available at: https://data.cms.gov/Medicare-Part-D/Medicare-Provider-Utilization-and-Payment-Data-Par/c9xz-7zpd [Accessed 11 Oct. 2018].