Predicting Over-Prescribers of Commonly Abused Medications

Milestone Report (09/2019)

Chika Nwugo

1 Introduction

On Jan 23rd, 2018, New York City filed a lawsuit against the manufacturers and distributors of commonly-abused prescription painkillers, with Mayor Bill de Blasio stating that "more New Yorkers have died from opioid overdoses than car crashes and homicides combined in recent years" [1].

The detrimental effects of prescription medications span far beyond NYC. According to the National Institute on Drug Abuse, over 70,200 Americans died from a drug overdose in 2017, which is almost double the number recorded in 2007 [2]. Opioids—mainly synthetic opioids (other than methadone)—are currently the main driver of drug overdose deaths in the U.S. [3].

Besides painkillers like opioids, Adderall, which is used to treat attention-deficit/hyperactive disorder (ADHD) a lesser-known but growing concern. The prescription of Adderall, which is known as the "study drug", is fast-growing and could catch up with that of opioids if nothing is done to check it. A reported 16 million prescriptions of Adderall were written in 2012, with sales quintupled from < $2 billion in 2002 to over $8 billion in 2012 [4].

Antibiotics are another frequently prescribed drug, despite over a quarter of doctors saying that antibiotics are often administered even though they will do no good, leading to waste in the healthcare system [5].

1.1 Objective

The objectives of this project are:

- To explore a sample dataset containing anonymized physician identities and their monthly patient volumes and payments received for Adderall, antibiotics, and opioids.
- To generate a classification system for physicians based on their prescription patterns
- To generate time series forecasts of patient volumes and payments if the prescription patterns continue without intervention

In this Milestone Report, we will present a description of the data set (Section 2), the steps for data cleaning and wrangling (Section 3), develop a data story which includes exploratory data analysis (EDA) and statistical analysis (Section 4), and outline the next steps of this project (Section 5).

The Jupyter notebook with Python codes used for this report can be found here.

1.2 Significance

Addressing what has inherently become an over-prescription epidemic [6], is no small task. Nonetheless, developing a pattern recognition classification system that helps to identify physicians that are likely over-prescribing medications can go a long way in mitigating the problem. Such a system can help:

- Patients make better decisions when selecting health care providers
- Insurance companies to flag over-prescribers and withhold payment for over-prescribed drugs, which ultimately protects patients and saves money
- Pharmaceutical companies and regulatory agencies to be more proactive in monitoring drug sales and minimizing abuse

1.3 References

[1] https://www.wsj.com/articles/new-york-city-sues-companies-over-opioid-abuse-1516734990
[2]https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates
[3]https://www.cdc.gov/drugoverdose/data/statedeaths.html
[4]https://www.nytimes.com/2013/12/15/health/the-selling-of-attention-deficit-disorder.html?pagewanted=all&_r=2
[5] https://www.cbsnews.com/news/doctors-still-overprescribing-drugs-antibiotics/
[6] https://www.statnews.com/2019/04/02/overprescribed-americas-other-drug-problem/

2 Dataset

The dataset used in this project is a sample construct based on real data from a domain that requires paid access. The data contains prescription information for patient monthly volumes and payments received for Adderall, antibiotics and opioid prescriptions by 3010 anonymized physicians from Jan-2015 to Oct-2017. The link to the dataset is here.

The dataset consists of 42074 entries (rows) and 7 features (columns). The column headings are:

1. PhysicianID – anonymized unique identifier for physician
2. yrmo – Year and month of prescription
3. AdderallPatients – total number of patients prescribed Adderall that month
4. AdderallPayments – total payments received from prescribing Adderall that month
5. Antibiopatients - total number of patients prescribed antibiotics that month
6. AntibioPayments - total payments received from prescribing antibiotics that month
7. OpioidPatients - total number of patients prescribed opioids that month
8. OpioidPayments - total payments received from prescribing opioids that month

3 Data Cleaning and Wrangling

The purpose of the data cleaning and wrangling steps are:

1. To ensure that all features are of the correct data type
2. To ensure missing data are properly imputed
3. To create additional potentially useful features
4. To prepare the dataset for exploratory data analysis (EDA) and statistical analysis

3.1 Data Type Correction

During dataset importation, the *yrmo* column was parsed as a pandas datetime object, renamed as "Date", and set as the data frame index.

3.2 Missing Values Imputation

Since not all physicians prescribed all three drugs every month, it was expected that there would be null entries for patient volume and payment for the months that a given drug was not prescribed by a physician. Nonetheless, 3 months were identified for antibiotics that had patient entries without corresponding payment entries. These rows were found to be associated with the same physician id, "GLwAKSb708GXX52I".

A sliced portion of the dataset with only the prescription history of said physician showed revealed 11 rows of data with antibiotic monthly patient volumes of 1 or 2 for the months without missing payments but antibiotic patient volumes of 57, 72 and 86 for the months with missing payment entries. The significant difference between the patient volumes for the missing payment entries and the none missing payments entries suggests that these rows are missing at random and will need to be imputed rather than deleted.

Based on the limited antibiotic entries for this physician, the suggested imputation method for the missing payment values was to calculate the average payment per patient for antibiotics from the entire dataset and use that value to multiply the number of antibiotic patients.

3.3 Incorrect Data Entry Correction

An investigation of the minimum and maximum values in each column revealed a "0" minimum entry for opioid patients associated with physician id "wwyJwO6CL1dqIooh". This entry was accompanied by an opioid payment entry of $63451. A sliced portion of the dataset with only the prescription history of said physician showed that the physician only prescribed opioids with monthly patient volumes ranging from 2 to 145. The data also suggested a significant correlation between the monthly patient volumes and payments for the said physician. This was confirmed by a Pearson correlation coefficient of 0.85 between monthly payments and patient volumes.

Since there was a significant payment amount associated with the "0" patient volume entry, the "0" was considered to be an error and linear regression interpolation was the suggested method to use for imputation.

4 Data Story

4.1 Descriptive Summary Statistics of the Raw Dataset

After imputations (from section 3), only the mean monthly antibiotic payments marginally increased 21682 from 21666, confirming that the imputation process did not significantly change the data.

Table 4-1 Descriptive summary statistics of numerical variables

Parameter	Adderall		Antibiotics		Opioid	
	Patients	**Payments**	**Patients**	**Payments**	**Patients**	**Payments**
Mean	4.9	1828.66	6.38	21682.74	3.1	2980.24
STD*	9.25	22930.74	12.74	83809.79	6.95	64217.29
CoV**	1.88	12.54	2	3.87	2.24	21.55
Minimum	1	-11771	1	-9400	1	-10
Median	3	245	3	2908	2	122
Maximum	341	1732264	285	2423404	318	2823021

*Standard Deviation, **Coefficient of Variation

Hence, the initial summary statistics of the dataset showed that on average:

- physicians saw about 5 adderall patients and charged about 1828 per month
- physicians saw about 6 antibiotic patients and charged about 21682 per month
- physicians saw about 3 opioid patients and charged about 2980 per month

Monthly patient volumes and payments showed considerable variation. The coefficient of variation analysis showed that monthly payments for opioids had the largest variation with a coefficient of variation of 21.5 followed by monthly payments for adderall, which had a coefficient of variation of 12.54.

Interestingly, it appears physicians received higher payments per patient for prescribing antibiotics compared to adderall or opioid. Likewise, it appears physicians received higher payments per patient for prescribing opioids compared to adderall. This could be because antibiotic and opioid prescribing physicians might prefer to prescribe more expensive branded products or might be overprescribing drugs to repeat patients.

The negative minimum monthly payments observed for drug types are suggested to result from patients that are unable to pay and the healthcare provider had to absolve the cost.

4.2 Estimating Probability Distributions of Variables

By extracting parameters from observed data and using those parameters to generate random variates, it was determined that the probability distributions of log-transformed values of monthly patient volumes appeared to have exponential distributions while that of monthly payments appeared to follow normal distributions (Figure 4-2).

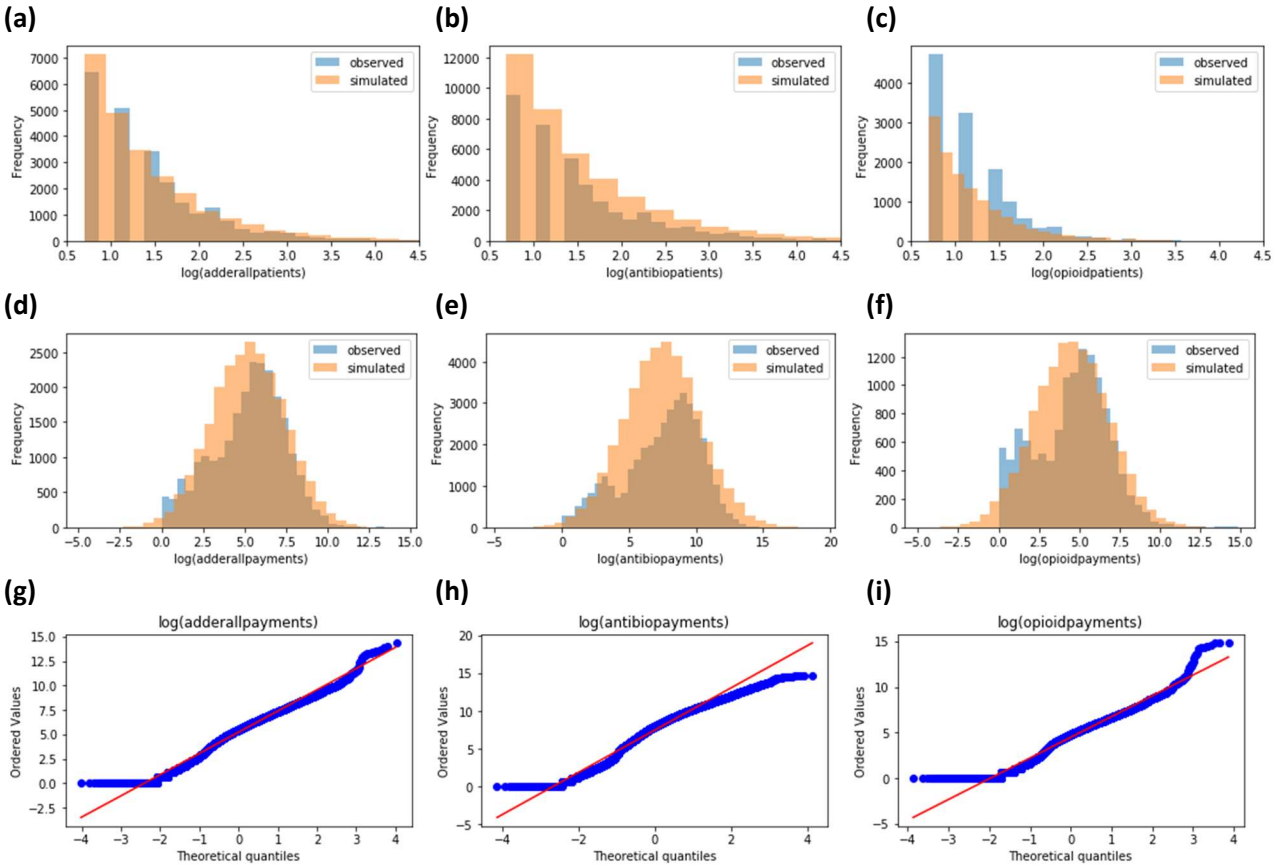

Figure 4-2 (a, b ,c) Histograms of observed and (exponential distribution) simulated values of log-transformed monthly patient volumes for adderall, antibiotics and opioids, respectively; **(d, e, f)** Histograms of observed and (normal distribution) simulated values of log-transformed monthly payments for adderall, antibiotics and opioids, respectively; **(g, h, i)** Quantile-quantile plots of log-transformed monthly payments for adderall, antibiotics and opioids, respectively.

4.3 Analysis of Monthly Patient Volumes and Payments

4.3.1 Box plots of Log Transformed Data

Irrespective of drug type, the minimum values, interquartile ranges and maximum values of patient volumes and payments suggest that they are right-skewed, prompting a log transformation. Box plots of

the log-transformed data still showed considerable right-skewness except for log-transformed monthly payments for antibiotics (Figure 4-1).

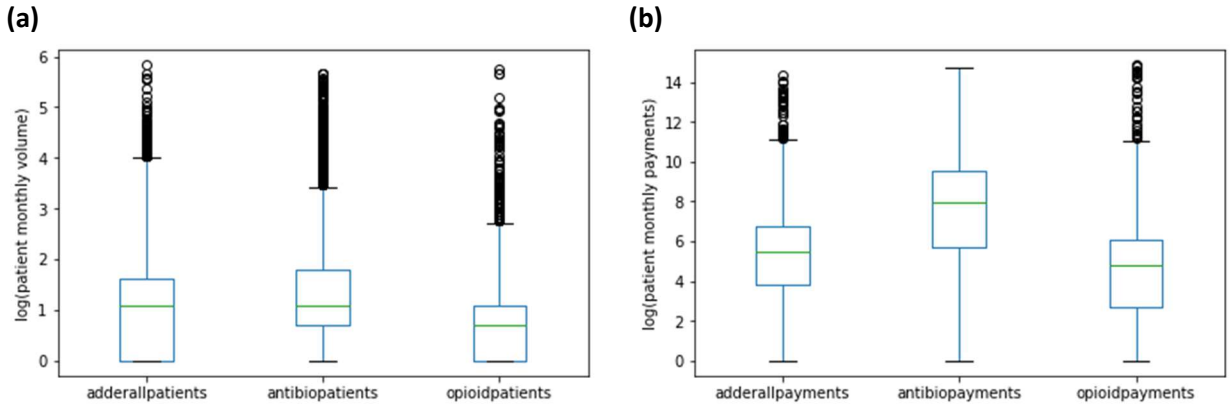

Figure 4-1 (a) Box plots of log-transformed values of monthly patient volumes for adderall, antibiotics and opioids; **(b)** Box plots of log-transformed values of monthly payments for adderall, antibiotics and opioids.

4.3.2 Empirical Cumulative Density Function (ECDF) of Log Transformed Data

The ecdf plots confirmed information from descriptive summary statistics (Table 4-1), which showed that antibiotics prescribing physicians had the highest monthly average patient volumes and highest monthly payments compared to adderall or opioid prescribing physicians. Additionally, ecdf plots showed that adderall prescribing physicians had higher monthly patient volumes compared to opioid prescribing physicians.

Interestingly, while descriptive summary statistics showed a higher monthly average payment for opioids compared to adderall, ecdf plots suggest that the 85th percentile or less of adderall prescribers received higher monthly payments compared to payments received by opioid prescribers. This suggests that the source of the higher average monthly payments for opioids compared to adderall is largely associated with payments above the 85th percentile for opioid prescribers (Figure 4-3). It is suspected that this group of opioid prescribers (> 85th percentile) might be involved in over-prescription practices.

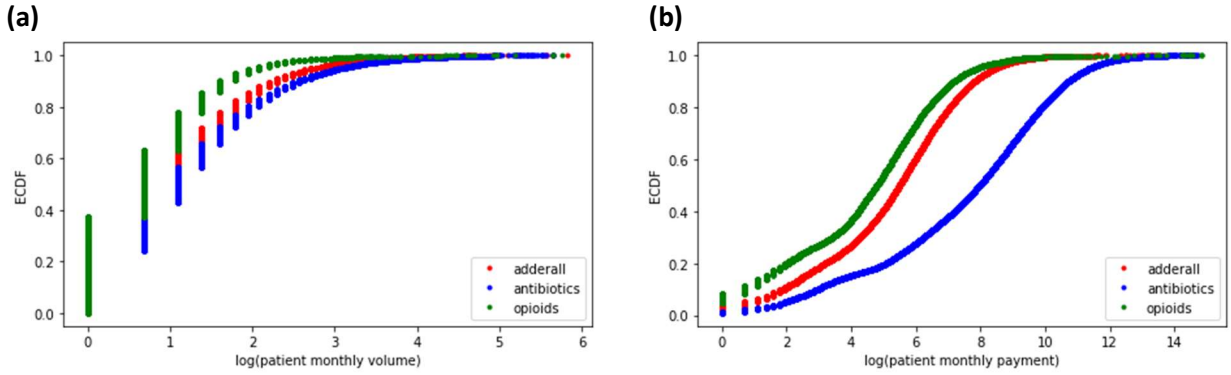

Figure 4-3 (a) Empirical cumulative density functions of log-transformed values of monthly patient volumes for adderall, antibiotics, and opioids; **(b)** Empirical cumulative density functions of log-transformed values of monthly payments for adderall, antibiotics and opioids.

4.3.3 Comparing Means of Monthly Patient Volumes and Payments

While summary descriptive statistics (Table 4-1) and ecdfs (Figure 4-3), clearly showed that on average, antibiotic prescribers saw the greatest volume of monthly patients and received the largest monthly payments, the inference for adderall and opioid prescribers were not as clear. For instance, descriptive summary statistics suggest that despite seeing fewer patients per month, on average, opioid prescribers received higher monthly payments than adderall prescribers. However, ecdfs of log-transformed data suggest that except for a few outlier opioid prescribers, adderall prescribers received larger monthly payments than opioid prescribers. Hence, to resolve this, it was important to test for a significant difference in mean monthly patient volumes and payments for adderall and opioids.

Bootstrap-assisted hypothesis test showed that the average monthly patient volume for adderall prescribers was significantly higher than that of opioid prescribers (*p = 0.0*). However, opioid prescribers had a significantly higher average monthly payment (*p = 0.0327*) compared to adderall prescribers at the 5% significance level.

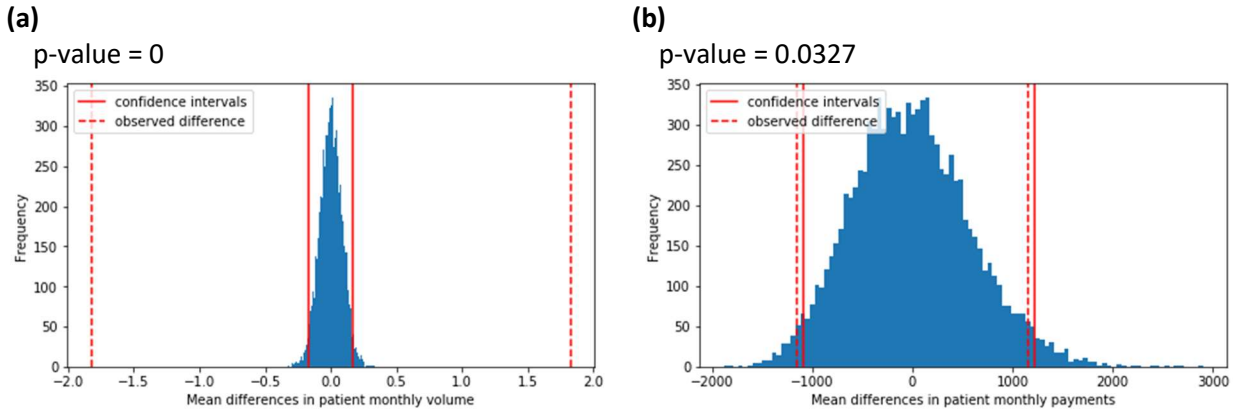

Figure 4-4 The p-values represent the probability of seeing the observed difference in group means if 10,000 randomized samples (with replacement) were compared from each group **(a)** Histograms of 10,000 bootstrap replicates of differences in means between monthly patient volumes for adderall and opioids; **(b)** Histograms of 10,000 bootstrap replicates of differences in means between monthly payments for adderall and opioids.

4.4 Analysis of Payment Per Patient (Payment Rate)

Since the data suggest a higher revenue margin for prescribing antibiotics and opioids compared to adderall, the average monthly payments received per physician per patient per dug was calculated. Hence, three new columns (variables) were created for the different drug types.

4.4.1 Descriptive Summary of Payment Rates

Descriptive summary statistics of payment rates showed that antibiotic prescribers had a significantly higher average monthly payment rate of $3017 compared to $259 for adderall prescribers and $250 for

opioid prescribers. The payment rates also showed considerable variation especially for opioids, which had a coefficient of variation of 5.15 compared to adderall and antibiotics with 3.21 and 2.86, respectively (Table 4-3).

Table 4-3 Descriptive summary of statistics of pay rates for adderall, antibiotics and opioids

Parameters	Adderall Payment Rate	Antibiotic Payment Rate	Opioid Payment Rate
Mean	259.68	3017.46	250.26
Standard Deviation	834.31	8636.15	1291.64
Coefficient of Variation	3.21	2.86	5.16
Minimum	-5885.50	-9400.00	-8.00
Median	85.09	887.00	62.50
Maximum	42753.00	649413.00	46170.35

4.4.2 Box plots and ECDFs of Payment Rates

Box plots and ecdfs show that physicians received significantly more revenue per patient by prescribing antibiotics. However, the average pay rate from opioid patients was only slightly higher than that for adderall patients.

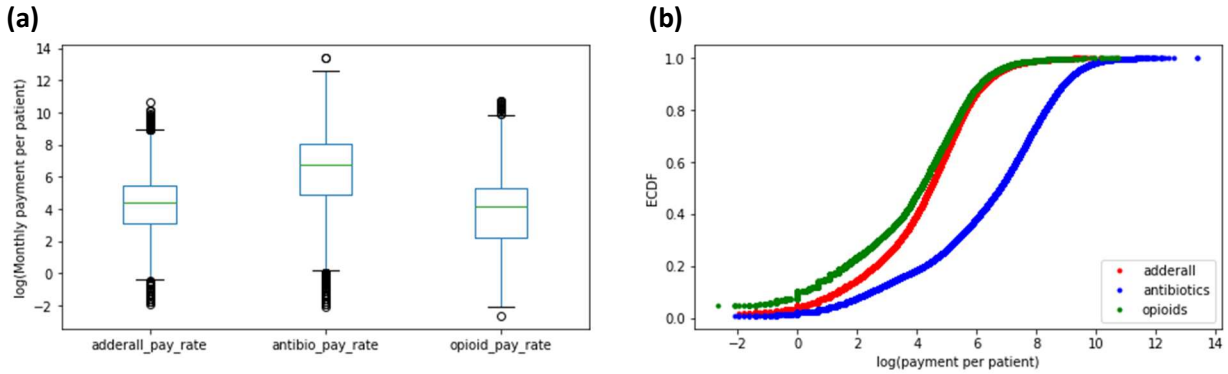

Figure 4-5 (a) Box plots of log-transformed values of monthly payments per patient for adderall, antibiotics, and opioids; **(b)** Empirical cumulative density functions of log-transformed values of monthly payments per patient for adderall, antibiotics, and opioids.

4.4.3 Comparing Means of Monthly Payments Per Patient

Bootstrap-assisted hypothesis test showed no significant differences in monthly payment per patient ($p = 0.7671$) at the 5% significance level (Figure 4-4).

Hence, it is interesting that there was no significant difference in average monthly payments per patient between adderall and opioid prescribers despite opioid prescribers having a significantly higher average monthly payment and a significantly lower average monthly patient volume compared to adderall prescribers.

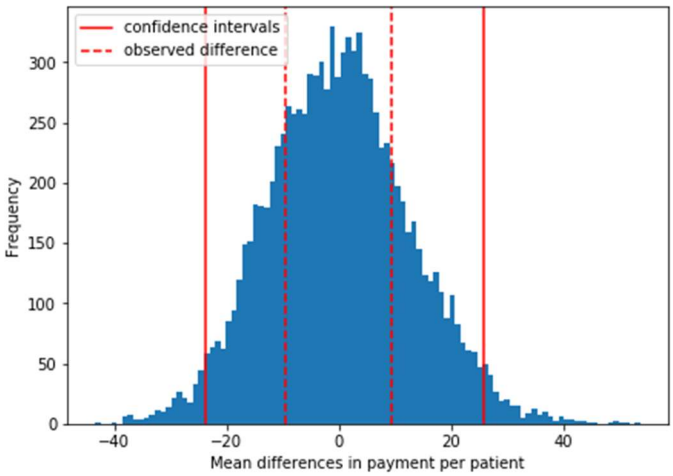

Figure 4-4 The p-value represents the probability of seeing the observed difference in group means if 10,000 randomized samples (with replacement) were compared from each group. Histograms of 10,000 bootstrap replicates of differences in means between monthly payments per patient for adderall and opioids.

4.5 Correlation Matrix of Variables

Correlation analysis showed a strong relationship between antibiotic patient volumes and antibiotic payments, which had a correlation coefficient (*r*) of 0.964 (Figure 4-5a and c). This suggests that despite having the highest monthly payment rate, antibiotic prescribers generally charged patients based on volume. Hence, one could easily predict the monthly payment to an antibiotic prescriber based on the monthly patient volume. This was not the case with adderall and opioid prescribers, which showed weak correlations between patient volumes and payments (Figure 4-5b and d). This suggests that other factors, besides patient volume, are responsible for the variation in monthly payments for adderall and opioids.

4.6 Detecting Over-prescribers

A function (***suspect_phys_finder***) was created to identify potential over prescribers of adderall, antibiotics, and opioids. Summary of the function's analytical steps are as follows:

1. For a given drug group, select prescribers that belong in the 80th percentile and above for monthly payments
2. For each prescriber selected from step 1, slice out the prescription history for the given drug
3. For prescribers with at least 3 active months, calculate the *r* between payments and patient volume
4. From the list of prescribers from step 1, select those with low *r* (i.e. those with r < 0.7 between payments and patients)

The results from running ***suspect_phys_finder*** are summarized in Table 4-4

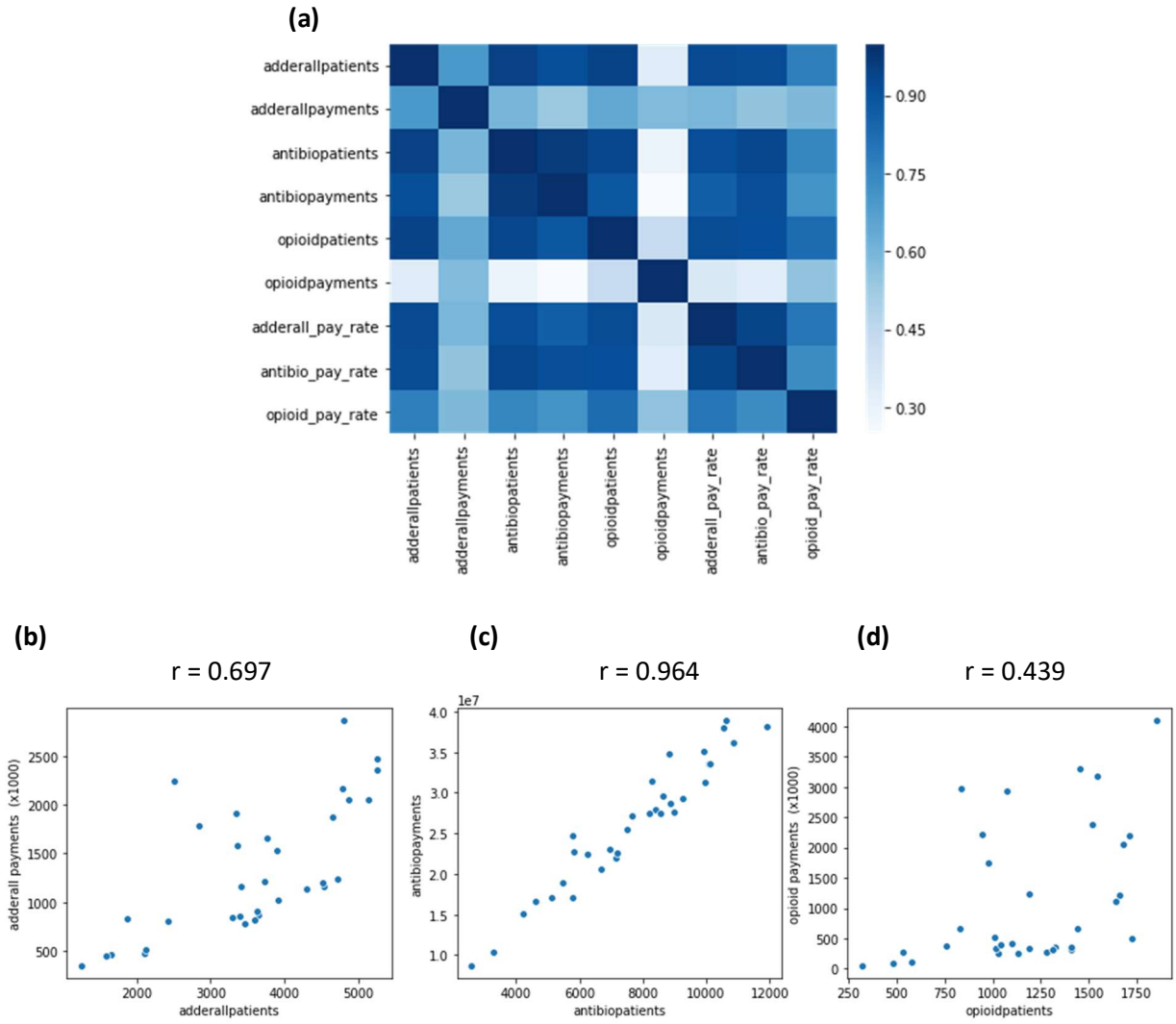

Figure 4-5 (a) Heatmap representation of the correlation matrix; **(b)** *r* and scatter plot of adderall payment *vs*. patients; **(c)** *r* and scatter plot of antibiotic payment *vs*. patients; **(d)** *r* and scatter plot of opioid payment *vs*. patients.

Table 4-4 Summary of number of potential over-prescribers within different prescriber groups

Prescriber Groups	Number of Over-prescribers	Number of Prescribers in Group
Adderall	143	2769
Antibiotics	118	2989
Opioids	139	2448
All drugs	0	2363
Adderall and antibiotics	10	2769
Adderall and opioids	16	2363
Opioids and antibiotics	14	2441

4.7 Time Series Analyses of Patient Volumes and Payments

The average monthly patient volumes, average monthly payments, and average payment rates did not show a trend and appeared stationary irrespective of drug type over the period (Figure 4-6a, b and c). For instance, the average monthly patient volume for antibiotic prescribers was approximately 6.1 in Jan 2015 and approximately 6.9 in Aug 2017. Additionally, the month of May 2017 had the highest average monthly patient volume of approximately 7.3 followed by Jun 2015 with approximately 7.1 for antibiotic prescribers (Figure 4-6a).

As expected, the average monthly patient volumes, average monthly payments and average monthly payment rates for antibiotic prescribers were consistently higher than those of adderall and opioid prescribers (Figure 4-6a, b and c). Subsequently, for patient volumes, the average monthly patient volumes for adderall prescribers were consistently higher than that of opioid prescribers (Figure 4-6a). However, for payments, the average monthly payments were about the same for adderall and opioid prescribers except for 10 months where the average payments were significantly higher for opioid prescribers compared to adderall prescribers (Figure 4-6b).

Interestingly, the total monthly patient volumes and payments showed a decreasing trend, especially for antibiotic prescribers (Figure 4-6d and e). This was later explained by the general decreasing number of prescribers (Figure 4-6f).

5 Conclusions and Next Steps

This report details the steps of data cleaning and wrangling as well as exploratory data analysis via visualizations and statistical methods.

Important insights generated from the analyses include:

- On average, antibiotic prescribers see more patients per month and generate more revenue compared to adderall and opioid prescribers
- While patient volume alone is a strong predictor of antibiotic prescription revenues, other factors, besides patient volume, appear to be predictors of adderall and opioid prescription revenues
- Adderall prescribers see more patients per month than opioid prescribers but generate less revenue
- The higher monthly revenues generated by opioid prescribers compared to adderall prescribers is most likely due to a small set of opioid over-prescribers
- Potential over-prescribers were identified using a function that selects prescribers above the 80th percentile for payments and with weak correlations between patient volumes and payments
- While the average monthly patient volume and average monthly payments were relatively stationary over the period, the total monthly patients and total monthly payments declined, which was attributed to a consistent decline in the number of prescribers

The next step of this project is to develop, evaluate and select machine learning models to classify prescribers based on their prescription profiles. This will be used to compare with prescribers identified

using the function. Time series forecasting will also be applied to predict the impact of the over-prescription of commonly abused medications.

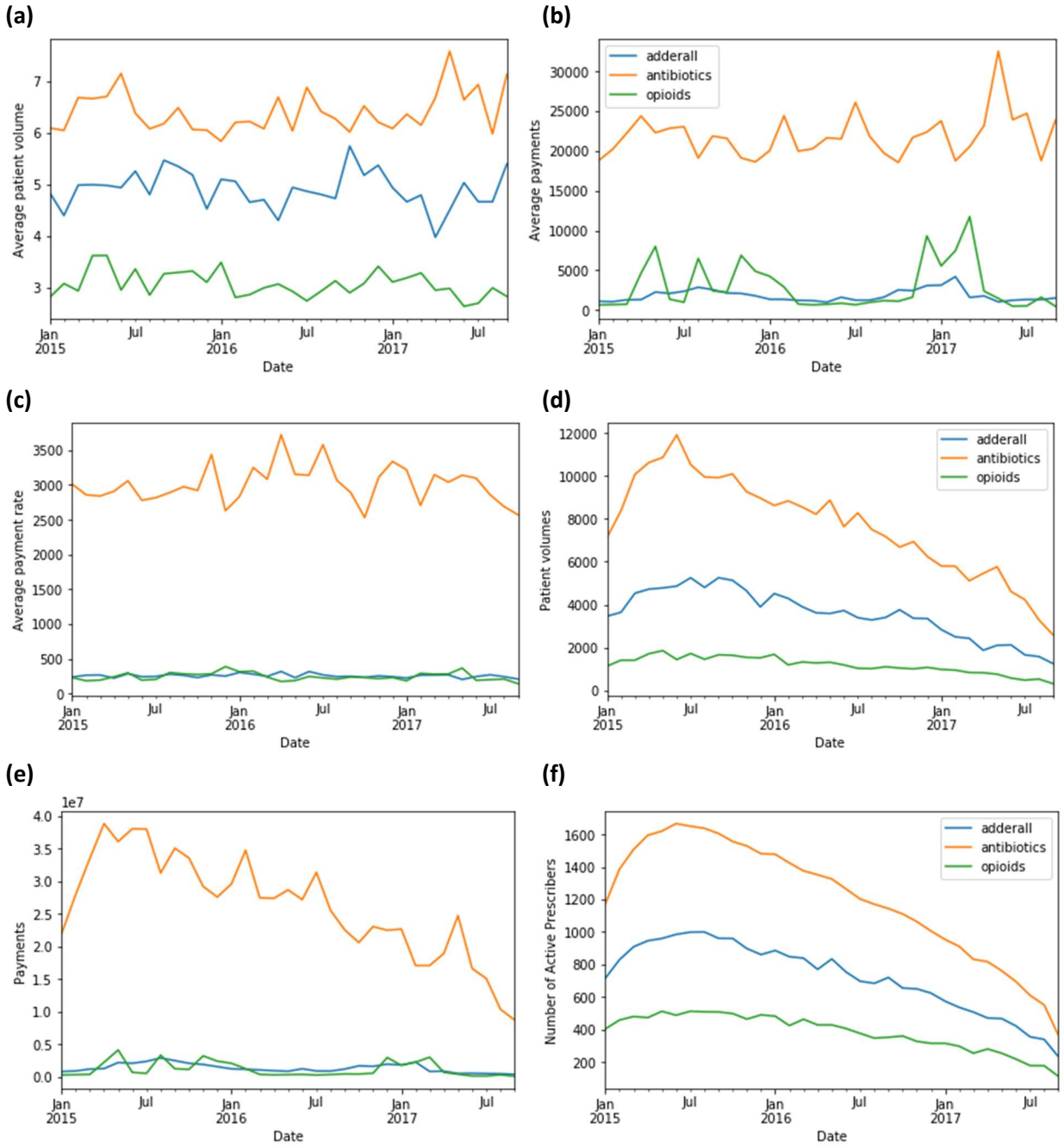

Figure 4-6 (a) Time series of average monthly patient volume; **(b)** Time series of average monthly payments; **(c)** Time series of average monthly payment rates; **(d)** Time series of total monthly patient volumes; **(e)** Time series of total monthly payments; **(f)** Time series of monthly counts of active prescribers.