Michelle Wang
Prof. Raman
ECON 255

Data Brief

Introduction

This report investigates a series of questions related to county characteristics, healthcare access, and opioid prescribing in the United States. The dataset that we are working with is a combination of several U.S. federal data sources containing county-level data on opioid prescriptions, county characteristics, and measures of healthcare access. The years captured in the dataset are 2006-2014 and our geographic subset is the state of California.

The county characteristic that I will be focusing on is veteran population estimate (F11396), and the measure of healthcare access that I will be exploring is the number of medical doctors per 100,000 residents (MD_PC). The questions that this report will investigate are:

- What is the relationship between the number of medical doctors per 100,000 residents and the veteran population estimate?
- What is the relationship between opioid prescribing and the number of medical doctors per 100,000 residents?
- What is the relationship between opioid prescribing and the veteran population estimate?

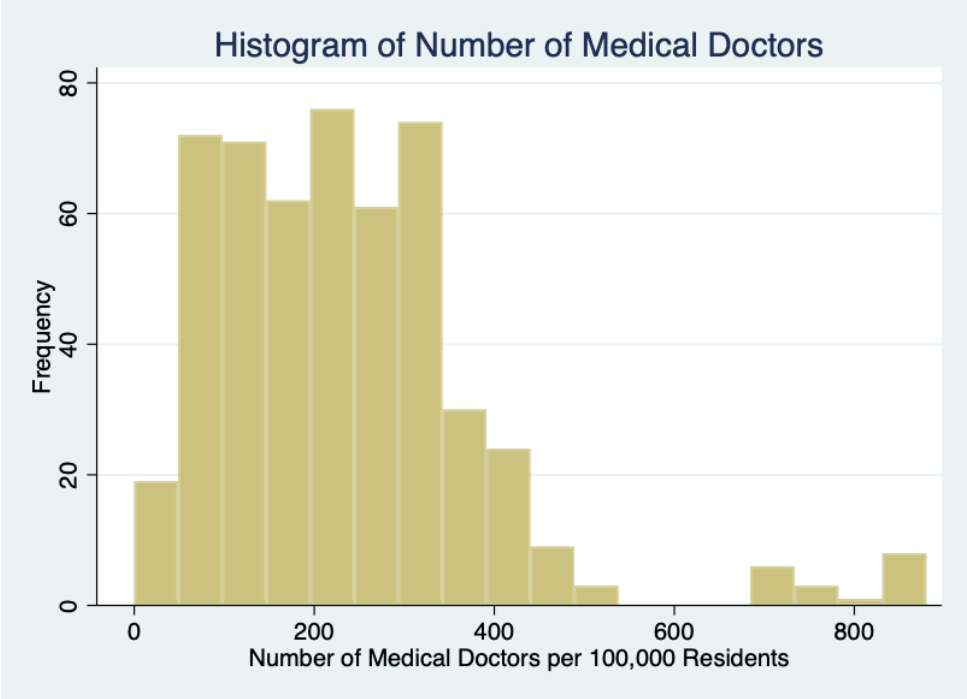
The number of medical doctors per 100,000 residents relates to opioid prescribing because having more medical doctors can increase the total amount of opioid prescriptions these medical doctors can collectively prescribe. The veteran population estimate relates to opioid prescribing because veterans are more likely to experience pain and stress due to physical, psychological, and social experiences before, during, and after military service, which they may use opioids to address, which affects opioid prescribing.

The specific variables used for analysis – the veteran population estimate and the number of medical doctors per 100,000 residents – are collected from the Health Resources & Services Administration's website. The opioid prescribing data is the monthly total number of opioid pills distributed, which was obtained from the U.S. Drug Enforcement Administration (DEA)'s Automation of Reports and Consolidated Orders System (ARCOS) pill shipment database, an extract of which was publicly-released by the Washington Post.

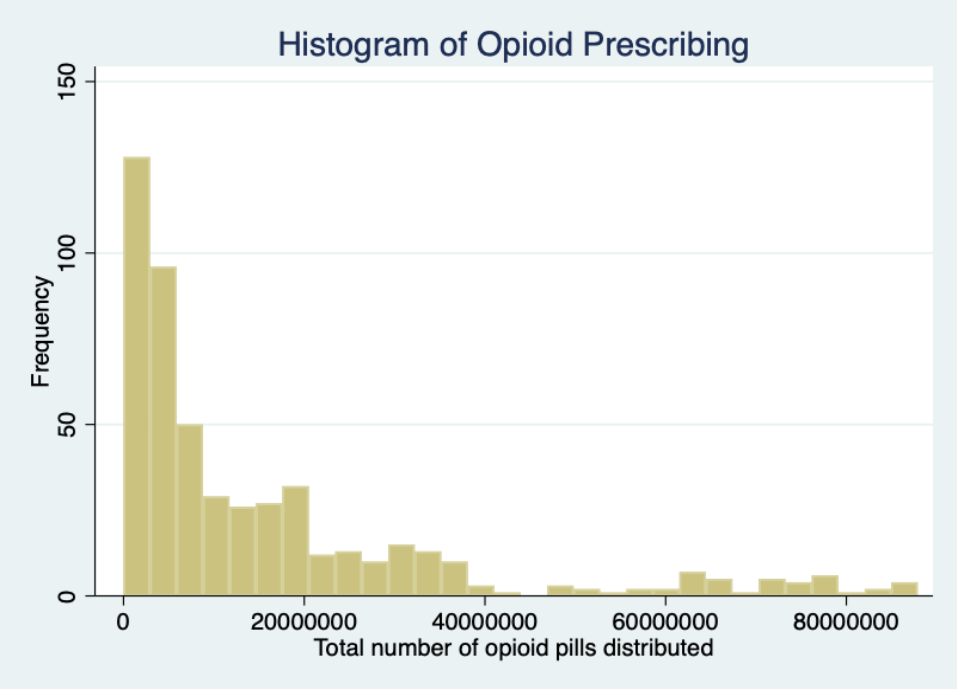
Some data hazards that this dataset may be associated with are Lacks community involvement, Reinforces existing bias, and Difficult to understand. The dataset may "lack community involvement" because it does not take in input from the communities represented in the dataset. It may "reinforce existing bias" such as the stereotype that low-income areas and

those in poverty are more likely to experience substance abuse and addiction. Finally, the dataset may be 'Difficult to understand' because many of the variables don't have representative names, such as the veteran population estimate we use for analysis in this report, which is represented as F11396. Although there is documentation to help identify variables, this still makes the dataset confusing and difficult to work with.

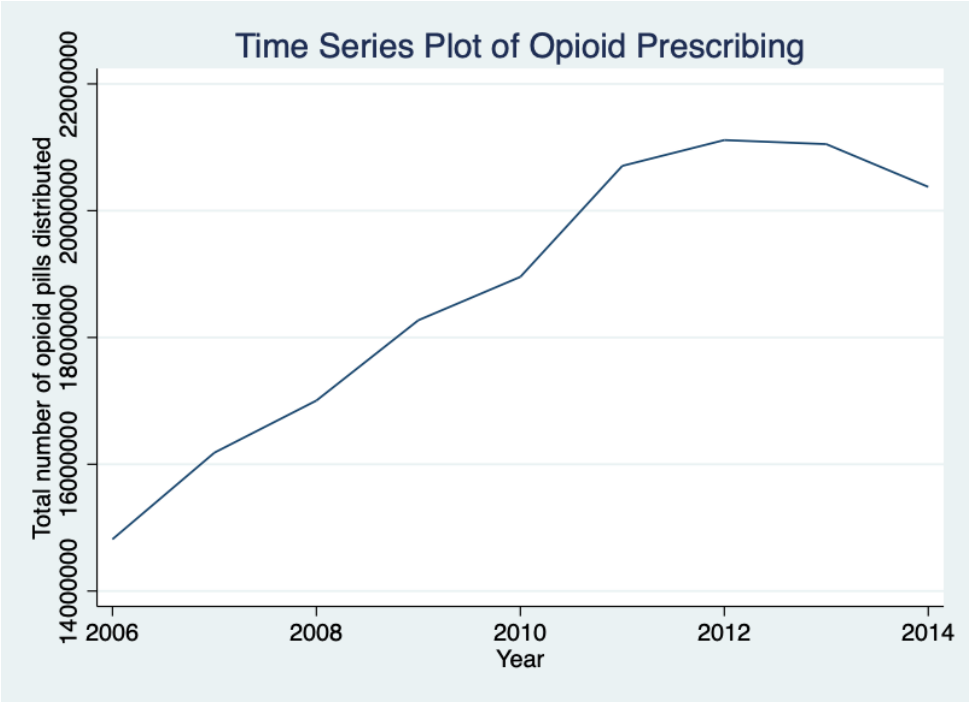
Descriptive statistics



This is a histogram showing the distribution of the number of medical doctors per 100,000 residents. The distribution is unimodal, slightly right-skewed, centered around 200, with some high outliers past 700.



This is a histogram showing the distribution of the monthly total number of opioid pills distributed, with outliers greater than 100,000,000 removed. The distribution is unimodal, extremely right-skewed, centered around 10,000,000, with some high outliers.



This is a time series plot of the monthly total number of opioid pills distributed over the years 2006 and 2014. We can see that the amount of opioid pills distributed increased during the years between 2006 to 2012, and then decreased slightly after 2012.

Econometric Analysis

	(1) DOSAGE_UNIT	(2) MD_PC	(3) DOSAGE_UNIT
MD_PC	43226.8***		
	(5.60)		
F11396		0.000584***	478.1***
		(8.05)	(20.98)
_cons	8467451.3***	217.1***	2322137.0***
	(5.58)	(28.86)	(4.60)
N	519	519	519

Discussion and Limitations

On average, for every one additional unit of the estimated veteran population, we predict about a 0.0005839 increase in the number of medical doctors per 100,000 residents. The p-value (0.000) is less than a significance level of $\alpha = 0.05$, which means that we have sufficient evidence to reject the null hypothesis that there is no relationship between the number of medical doctors per 100,000 residents and the estimated veteran population.

On average, for every one additional medical doctor per 100,000 residents, we predict about a 43226.78 increase in the monthly total number of opioid pills distributed. The p-value (0.000) is less than a significance level of $\alpha = 0.05$, which means that we have sufficient evidence to reject the null hypothesis that there is no relationship between the monthly total number of opioid pills distributed and the number of medical doctors per 100,000 residents.

On average, for every one additional unit of the estimated veteran population, we predict about a 478.105 increase in the monthly total number of opioid pills distributed. The p-value (0.000) is less than a significance level of $\alpha = 0.05$, which means that we have sufficient evidence to reject the null hypothesis that there is no relationship between the monthly total number of opioid pills distributed and the estimated veteran population.

However, our analysis is limited by the fact that we did not check to see if the four OLS assumptions – the relationships between each pair of variables is linear, observations are independent of each other (the sample is randomly assembled), the variance of the residual term is constant, and there are no large outliers – are satisfied. If the assumptions for OLS are not satisfied, and this negatively affects the reliability and validity of our linear regression estimates. Additionally, my chosen geographic subset – which includes only the counties in California – excludes data from all other US states, which means that we cannot generalize our analysis results on a national level.