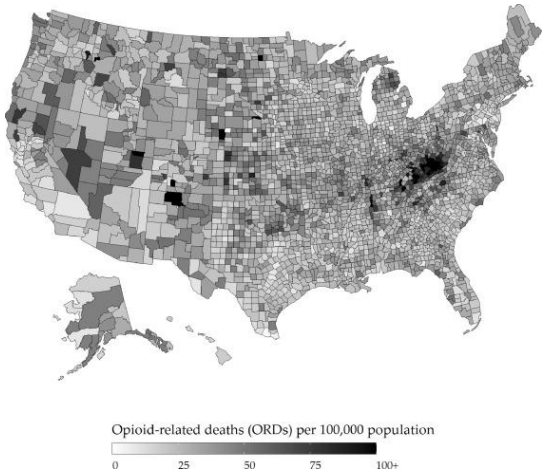
Econometrics Data Brief #1

Introduction

The data analyzed in this report is a combination of several U.S. federal data for the years 2006-2013. The motivation for creating this dataset originated from the want to assist researchers in identifying which factors predict per capita opioid pill volume (PCPV) in a county, and whether early state Medicaid expansions increased PCPV and PCPV's association with opioid-related mortality. The data contains raw data from federal websites and preprocessed data regarding opioid volumes. Additionally, the dataset is organized by county level and contains healthcare and demographic data for opioid pill volumes on the county level.

With any sample dataset, there may be some bias. One notable area for potential bias within the dataset is that this dataset is from 2006-2014. It is likely that the opioid epidemic has changed over time, and thus we must interpret our results carefully. Additionally, it is likely that policy relating to the opioid epidemic has changed over time, so even if the data shows significant results, there may not be any use for our findings. Another data hazard for this data set is that the majority of the data for this dataset was taken from U.S. federal data. The U.S. federal data is largely collected from surveys and thus can lead to selection bias. Perhaps certain hospitals or counties contributed more data compared to other counties. Additionally, it is mentioned that there are missing data points, so we must be cautious of this as well.

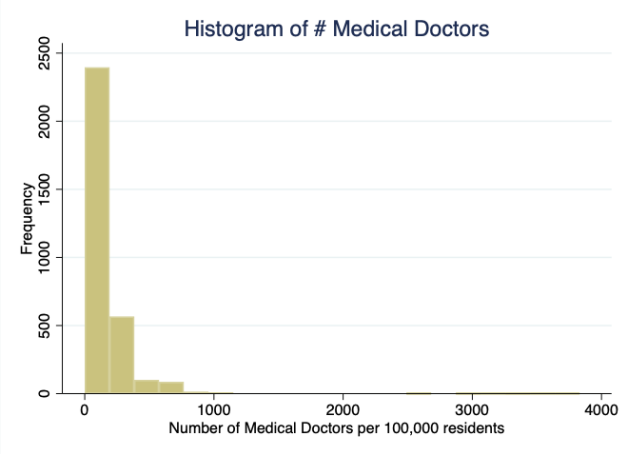


When initially looking at the dataset description, Figure 2 shows an interesting pattern at first glance regarding mean annual opioid-related deaths by county[1]. The figure shows that the Appalachia area displays, on average, a highly concentrated death count for opioid-related deaths. Thus, the scope of this analysis will be limited to the Appalachia region, specifically the states as follows: Ohio, West Virginia, Kentucky, Pennsylvania, and Maryland. The measure of healthcare access explored in this analysis is the number of medical doctors per 1000,000 county residents ("MD_PC"). From an initial macroscopic read, this quantity seems like a good indicator of the level of healthcare access in a county. Additionally, one would guess poverty would be a main driver of opioid usage and opioid-related deaths. However, based on the graph, it is known that there are broad geographic patterns other than poverty that could be indicators as poverty exists in a wide variety of regions in the United States. The county characteristic that will be investigated in relation to opioid-related county deaths is the unemployment rate for 16+ (F06795).

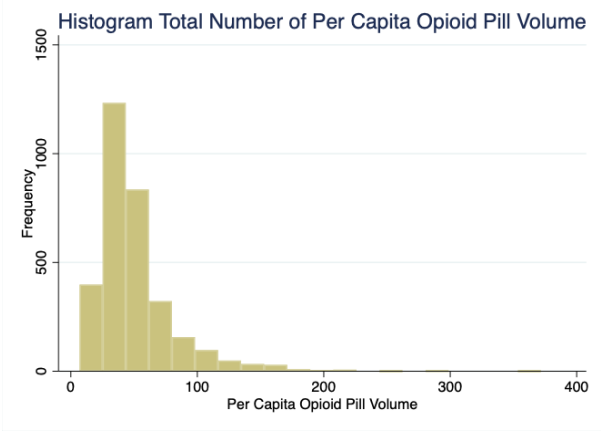
[1] N. Griffith, Kevin, et al. "County-Level Data on U.S. Opioid Distributions, Demographics, Healthcare Supply, and Healthcare Access." *Data in Brief*, Elsevier, 30 Jan. 2021, www.sciencedirect.com/science/article/pii/S2352340921000639.

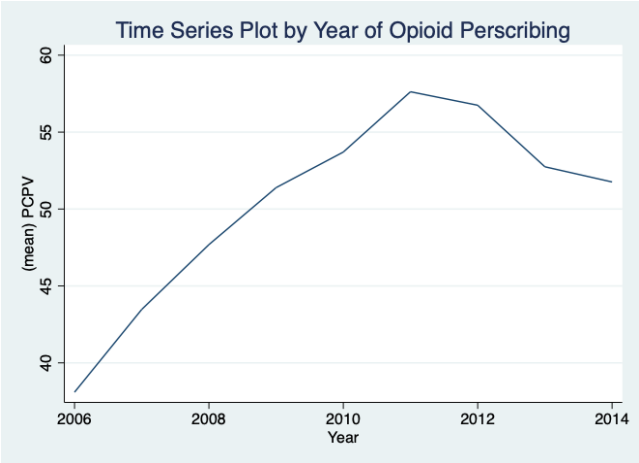
Descriptive Statistics

The measure of healthcare access chosen for this analysis is the number of medical doctors per 100,000 residents (represented as variable "MD_PC" in the data). MD_PC is a good measure of healthcare access because ultimately, to receive healthcare or medical assistance, there needs to be medical doctors. Additionally, opioids are only to be legally prescribed through healthcare professionals. While this variable, only includes medical doctors, it is interesting to study the distribution of what is considered one of the highest healthcare positions with the other opioid data in this study. The histogram to the right shows that the distribution of medical doctors in the Appalachian region is right-skewed, with the majority of observations concentrated around 20 medical doctors per 100,000 residents. There are extreme values with low frequency or high outliers, causing the distribution to be skewed.



To measure opioid prescribing, this analysis looks at the opioid pill volume per capita. The distribution of the histogram for the total number of per capita opioid pill volume is also right skewed, with most of its values concentrated at 40 opioid pill volume per capita. Additionally, there are a few high outliers, indicating that some counties in the Appalachian region have high opioid pill volume per capita.





To the left is a time series plot from the years 2006-2014 of the average opioid pill volume per capita in the Appalachian region. The plot shows that opioid pill volume increased from 2006-2011, and then decreased from 2011-2014. The plot indicates that there could've been environmental changes impacting the Appalachian region in 2011 resulting in changed opioid pill volumes.

Econometric Analysis

The interpretation for the linear regression of the number of medical doctors and the unemployment rate for 16+ can be explained as follows (Column 1). A 1 percent increase in the

unemployment rate for 16+ is associated with a decrease of 18.18874 medical doctors per 100,000 county residents in the Appalachian region. Additionally, when the unemployment rate for 16+ is 0%, the expected number of medical doctors per 100,000 county residents is 307.1122 in the Appalachian region. It is reasonable that the number of medical doctors is very high compared to the mean because it is impossible for the unemployment rate to ever be as low as 0% due to macroeconomic factors like frictional unemployment. The null hypothesis test for this regression tests if the coefficient for the unemployment rate of 16+ is equal to 0, and thus has no effect on the # of medical doctors per 100,000 county residents. The alternative hypothesis is that there is an association between the two variables, or that the coefficient is not equal to 0. The p-value for this hypothesis test is equal to 0. Therefore we reject the null hypothesis test and have statistical evidence that there is not no association between the unemployment rate and the number of medical doctors in a county.

Table of Regression Outputs

	(1) MD_PC	(2) PCPV	(3) PCPV
F06795	-18.19***		3.482***
	(-11.95)		(15.04)
MD_PC		-0.00432**	
		(-2.90)	
_cons	307.1***	51.07***	22.95***
	(20.42)	(82.61)	(13.30)
N	3177	3177	3177

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Column 1: # of Medical Doctors per 100,000 (MD_PC) vs. Unemployment Rate for 16+ (F06795)
Column 2: Opioid Prescribing per Capita Opioid Pill Volume(PCPV) vs. # of Medical Doctors per 100,000 (MD_PC)
Column 3: Opioid Prescribing per Capita Opioid Pill Volume(PCPV) vs. Unemployment Rate for 16+ (F06795)

Second, we can analyze the regression between the per capita pill volume (PCPV) and the # of medical doctors per 100,000 county residents (MD_PC) in the Appalachian region. The regression model shows that for an increase in 1 medical doctor per 100,000 county residents there is an associated decrease in per capita pill volume of 0.00432 on average. Additionally, when the number of medical doctors per 100,000 county residents is 0, the estimated per capita pill volume is 51.07. 51.07 is similar to the mean for the distribution of PCPV in the Appalachian region. The null hypothesis test for this regression tests if the coefficient for the # of medical doctors is equal to 0, and thus has no effect on the per capita pill volume. The alternative hypothesis is that there is an association between the two variables, or that the coefficient is not equal to 0. The p-value for this hypothesis test is equal to 0. Therefore we reject the null hypothesis test and have statistical evidence that there is not no association between the number of medical doctors in a county and per capita pill volume in the Appalachian region.

Third, we can analyze the regression model between the per capita pill volume (PCPV) and the unemployment rate for 16+ (F06795) in the Appalachian region. The model predicts that a 1 percent increase in the unemployment rate for 16+ is associated with a per capita opioid pill volume increase of 3.482 in the Appalachian region. Additionally, when the unemployment rate

for a county is 0% in the Appalachian region, it is estimated that the per capita pill volume is 22.95 units. The null hypothesis test for this regression tests if the coefficient for the unemployment rate of 16+ is equal to 0, and thus has no effect on the per capita pill volume. The alternative hypothesis is that there is an association between the two variables, or that the coefficient is not equal to 0. The p-value for this hypothesis test is equal to 0. Therefore we reject the null hypothesis test and have statistical evidence that there is not no association between unemployment rate and per capita pill volume in the Appalachian region.

Discussion and Limitations

Based on the linear regression, there seems to be a correlation between each explanatory variable and the corresponding dependent variable. For the first regression, there is an inverse relationship between the unemployment rate for 16+ and the number of medical doctors per 100,000 county residents in the Appalachian region. This correlation is coherent because unemployment rate can be an indicator of poverty, meaning that when unemployment rates increase, poverty levels increase. High levels of poverty in an area often result in households being unable to afford medical care (especially because many jobs provide medical insurance). Consequently, medical doctors may choose not to practice in these areas as they may not be able to make a living. The second linear regression model shows that there is a positive association between a number of medical doctors in an area and the per capita opioid pill volume in the Appalachian region. This relationship also is sensible because medical doctors are part of the group of people legally able to provide opioid prescriptions. It makes sense, that when there are more doctors in an area, there is an expected increase in opioid pill volume. The regression model between unemployment rate and per capita pill volume shows that there is a suggested positive association between the two variables in this region. Therefore both unemployment rates and the number of medical doctors per capita have a positive association with opioid pill volume per capita. However, the unemployment rates for 16+ variable has a negative association with the number of medical doctors per 100,000 residents.

Many limitations exist in this analysis. The first limitation arises from the definition of the measure of healthcare access. In this analysis, the measure of healthcare access is defined as the # of medical doctors per 100,000 residents. However, healthcare access is much more extensively defined and larger in scope than the # of medical doctors in an area. In the context of this data, it was whether varying levels of healthcare access impact the amount of opioids in a given county. One example of how this limitation could impact our findings is that defining healthcare access as just the # of medical doctors leaves out nurses and physician assistants. Depending on the state, nurses and physician assistants are able to prescribe opioids as well, so leaving this data out of our regression models could've caused significantly different results. Another limitation of this analysis is that the data is aggregated by county. Aggregating by county can lead to generalization in our results. For example, rural and urban areas that are in the same county may have varying levels of income and access to healthcare. However, these varying relationships would not be captured in our model.

Lastly, I would argue that the biggest limitation of this analysis is that it is only concentrated on the Appalachian area. It is noteworthy to study potential opioid consumption factors in this region because of the region's high level of opioid existence. However, without comparing these factors or possible predictors in areas where little to know opioid pill volume exists, it can be difficult to conclude significant findings. Additionally, the regression models create predicted associations based on the entire Appalachian area, when it is likely that there is variability across states and counties in the region.