Milestone 7

Ali Crump

4/16/2020

Abstract

Zoorob (2019) shows that geography and fentanyl exposure explain much of the variation in increased overdose mortality rates between 2011 and 2017. This paper was mostly successful in replicating the results of the original paper, however there are discrepancies in the fentanyl exposure estimates and the total estimated deaths attributable to fentanyl for each model. Some of the estimates in this replication are 13% larger than those published, and the author has been notified and is working on corrections. In addition to replicating Zoorob's work, the extension of this paper aims to adjust the definition of fentanyl exposure using log transformations and to display a range of uncertainty in the total estimated deaths. The focus of the extension is on Zoorob's ordinary least squares model and I find X. This is important because Y.

Introduction

Zoorob uses two models; Model 1 shows that fentanyl exposure has a positive association with mortality rates, and Model 2 tries to estimate the causal effect of fentanyl exposure on mortality rates. Zoorob runs a least squares regression for the first model. The model predicts overdose mortality as a function of fentanyl exposure. Fentanyl exposure takes into account the state, year, an error term, and the natural logarithm of the number of test results containing fentanyl:

$Fentanyl_{ij} = \log{(\frac{S_{ij}}{P_{ij}} + 1)}$

Model 1 below is an ordinary least squares equation where α_i is state i and η_j is year j The standard errors are two-way clustered by state and year and includes population weights (@Paper).

$Overdose_{ij} = \alpha_i + \eta_j + \beta_1 Fentanyl_{ij} + \epsilon_{ij}$

The second model uses a two-stage least squares regression:

$\widehat{Fentanyl}_{ij} = \alpha_i + \eta_j + \beta_1(Longitude_i \cdot Year_j) + \epsilon_{ij} Overdose_{ij} = \alpha_i + \eta_j + \beta_2 \widehat{Fentanyl}_{ij} + \epsilon_{ij}$

Findings in the paper show that much of the variation in the increased overdose mortality is explained by fentanyl exposure, and that fentanyl deaths are highly correlated with geography, as the epicenter of the overdose crisis has shifted towards the eastern U.S. They also found that longitude is better able to explain levels of overdose mortality over time. States east of the Mississippi River tend to have greater fentanyl exposure and sharper increases in overdose deaths than states west of the Mississippi River (@Paper). Zoorob also uses both models to estimate the number or overdose deaths attributable to fentanyl and claims that they are broadly consistent with official mortality statistics.

Zoorob obtained the data used for his analysis through a Freedom of Information Act request. The data consist of state test results for drug seizures between 2011 and 2016, which he filters for test results containing fentanyl. Zoorob also uses age-adjusted mortality data from the National Center for Health Statistics. All the data used contain state and year information, and he uses state-annual populatons to calcaulte mortality rates relative to a state's population in a particular year. The data and code that Zoorob used in his paper is available on the Harvard Dataverse. To conduct my replication, I used R. More information on this project can be found on my Github repository.[1]

What did I do?

Redefine $Fentanyl_{ij} = \log{(\frac{S_{ij}}{P_{ij}} + 1)}$ Display range of uncertainty

What did I find?

Zoorob's method for measuring fentanyl exposure is the closest to official mortality statistics. What is the range of uncertainty

Literature Review

relevant literature in the paper

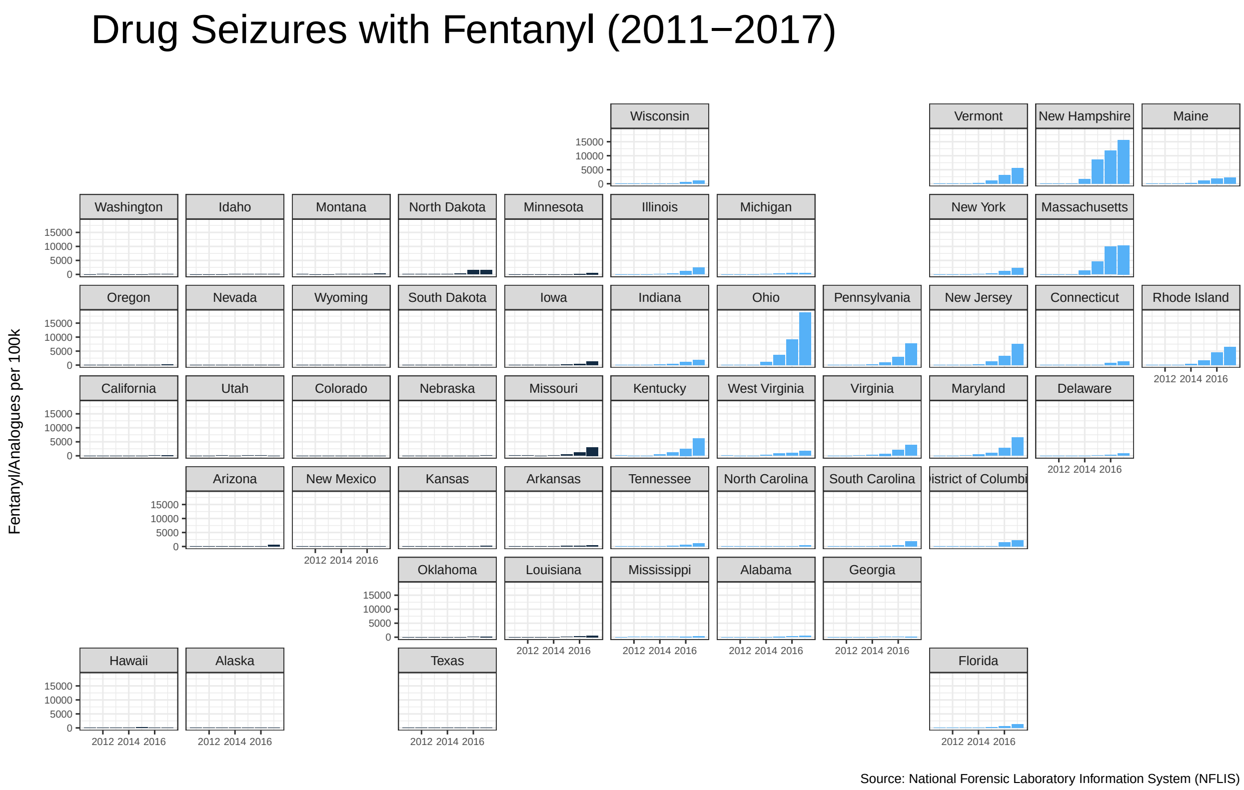
The number of drug overdose deaths in the United states has rapidly increased since 2014. However, the opiod epidemic did not affect all regions of the U.S. equally; according to the CDC, almost all states west of the Mississippi River did not see an increase while those to the west did. While Dasgupta et. al argue social and economic factors play a role in one's susceptibility to opiod addiction and overdose, Zoorob claims that the geographical patterns point to drug supply also playing a primary role in the epidemic.

relevant literature since the paper published

Research by Barocas et al. on the effect of opiod use among patients with endocarditis cites Zoorob's paper and also finds a decreased risk of overdose associated with the West and South compared to the Northeast (@endocarditis). Although in 2018, drug overdose deaths actually decreased by 4.6% from 2017 in the United States, fentanyl deaths continued to rise (@CDC)(@NYT). (Still looking for literature on this)

[1] Github repository

Appendix



tables 2a 2b

table 2 fentr = model 1 fent_r(fit) = model2

Table 1:

	Dependent variable:	
	age_adjusted_rate	
	(1)	(2)
fent_r	4.508***	
	(0.635)	
‘fent_r(fit)‘		5.443***
		(0.653)
Observations	357	357
R^2	0.928	0.923
Adjusted R^2	0.914	0.908
Residual Std. Error (df = 299)	5,372.861	5,545.678
Note:	*p<0.1; **p<0.05; ***p<0.01	

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Wed Apr 22 11:38:14 2020

supplementary table a1 in appendix same as table above but omitting alaska and hawaii

	Model 1 Deaths	Model 2 Deaths
2011	2580	3115
2012	2659	3210
2013	3723	4495
2014	9973	12041
2015	17367	20969
2016	26491	31985
2017	34176	41263

Table 2:

	Dependent variable:	
	age_adjusted_rate	
	(1)	(2)
fent_r	4.546***	
	(0.644)	
‘fent_r(fit)‘		5.573***
		(0.721)
Observations	343	343
R^2	0.928	0.923
Adjusted R^2	0.915	0.908
Residual Std. Error (df = 287)	5,455.543	5,665.766
Note:	*p<0.1; **p<0.05; ***p<0.01	

table 1 and correlates

longitude r squared

```
## [1] "2011: 0.000133604247072976"
## [1] "2011: (NO Hawaii/Alaska) 0.0204213094278517"
## [1] "2012: 0.00170481159966044"
## [1] "2012: (NO Hawaii/Alaska) 0.0273578839355274"
## [1] "2013: 0.0240861774892379"
## [1] "2013: (NO Hawaii/Alaska) 0.00464511710739798"
## [1] "2014: 0.291779582731855"
## [1] "2014: (NO Hawaii/Alaska) 0.339755720111154"
## [1] "2015: 0.324391935198356"
## [1] "2015: (NO Hawaii/Alaska) 0.447996991934456"
## [1] "2016: 0.463180674808103"
## [1] "2016: (NO Hawaii/Alaska) 0.538614954758827"
## [1] "2017: 0.554045298467256"
## [1] "2017: (NO Hawaii/Alaska) 0.573811061894831"
```

latitude r squared

```
## [1] "2011: 0.022522919018938"
## [1] "2011: (NO Hawaii/Alaska) 0.0229708939591238"
## [1] "2012: 0.0403295747755529"
## [1] "2012: (NO Hawaii/Alaska) 0.0265512094159112"
```

```
## [1] "2013: 0.0404112441498816"
## [1] "2013: (NO Hawaii/Alaska) 0.0378140648011715"
## [1] "2014: 0.0250983739834497"
## [1] "2014: (NO Hawaii/Alaska) 0.0293815482091747"
## [1] "2015: 0.00204162860211"
## [1] "2015: (NO Hawaii/Alaska) 0.0285509634034103"
## [1] "2016: 0.0315754568568326"
## [1] "2016: (NO Hawaii/Alaska) 0.0418356152848179"
## [1] "2017: 0.0103733848766385"
## [1] "2017: (NO Hawaii/Alaska) 0.0206704172153174"
```

mortality 2013 r squared

```
## [1] "2011: 0.0285374370715816"
## [1] "2011: (NO Hawaii/Alaska) 0.0229136041040015"
## [1] "2012: 0.00512744385520949"
## [1] "2012: (NO Hawaii/Alaska) 0.00875256024148491"
## [1] "2013: 0.00643192087668385"
## [1] "2013: (NO Hawaii/Alaska) 0.0107743144223076"
## [1] "2014: 0.0967962132363055"
## [1] "2014: (NO Hawaii/Alaska) 0.089162721462024"
## [1] "2015: 0.0540464695899684"
## [1] "2015: (NO Hawaii/Alaska) 0.0546523015095623"
## [1] "2016: 0.0384653767725879"
## [1] "2016: (NO Hawaii/Alaska) 0.0317295820351081"
## [1] "2017: 0.0295998137577852"
## [1] "2017: (NO Hawaii/Alaska) 0.0226044358364418"
```

table 1

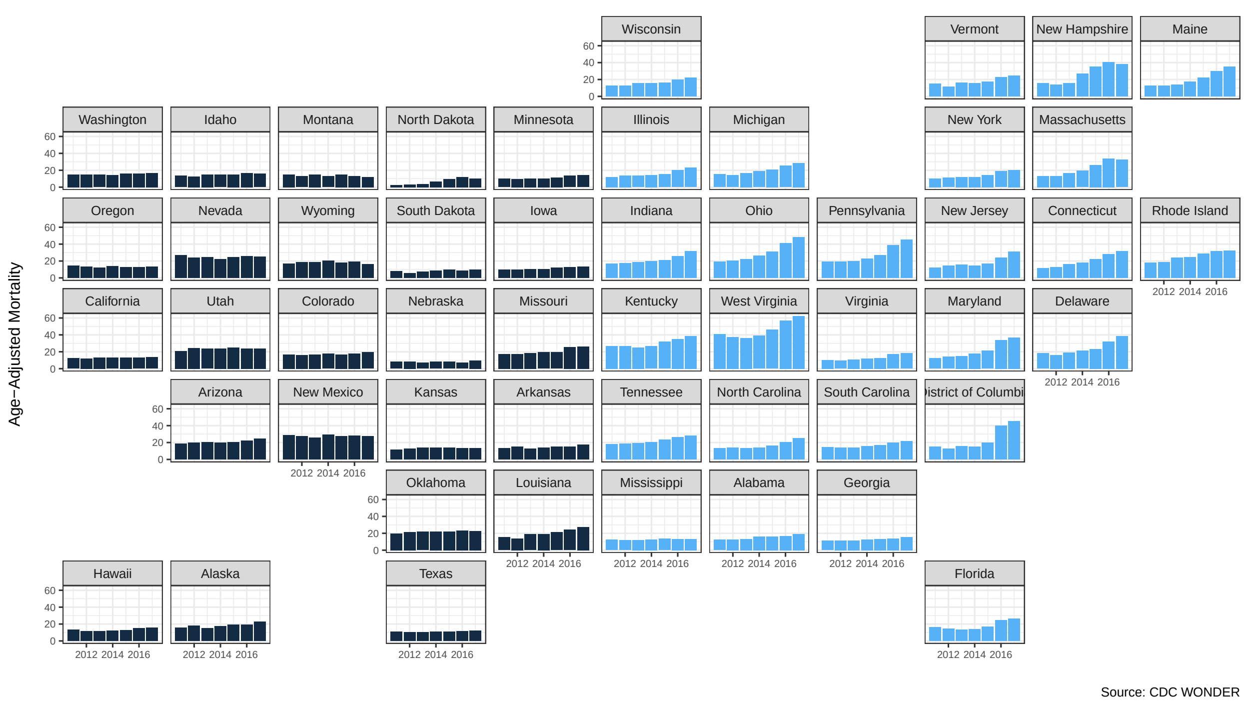
begin figure 2a and figure 2b paper

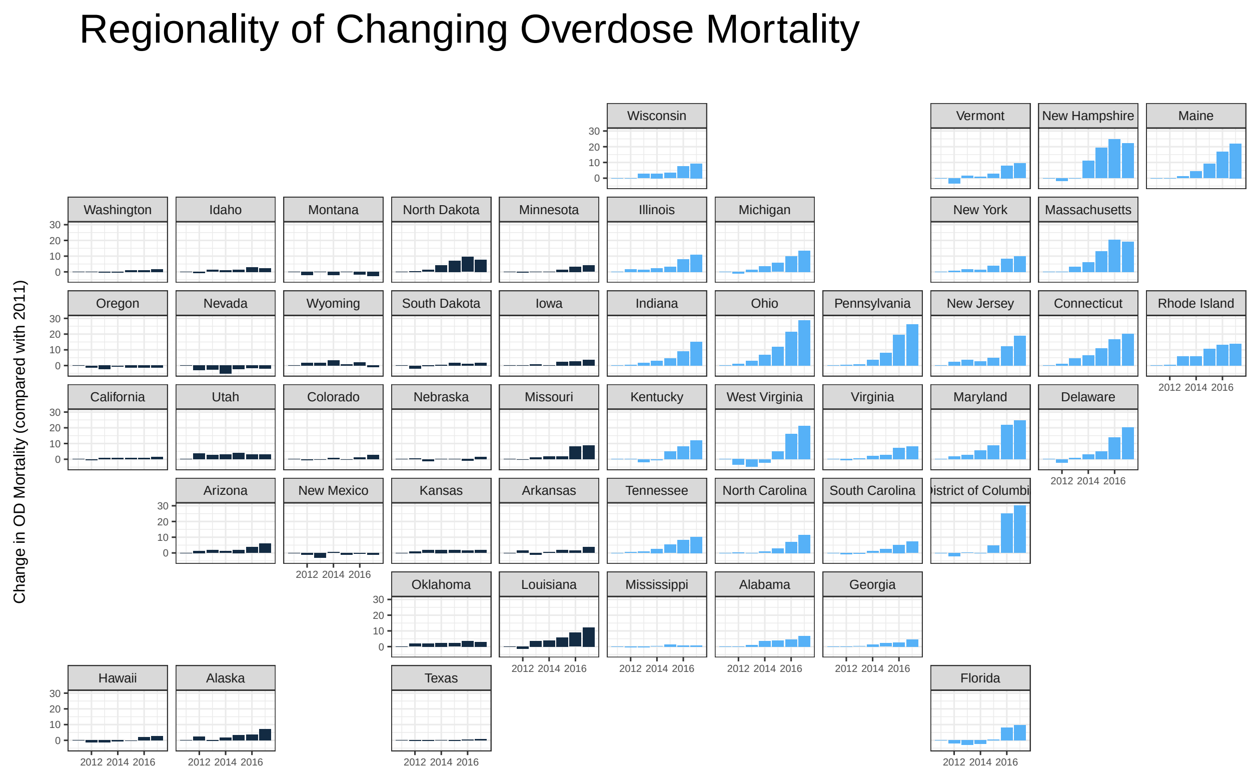
Table 3:

	Dependent variable:				
	fent_r				
	(1)	(2)	(3)	(4)	(5)
longitude	0.002	0.019***	0.030***	0.044***	0.053***
	(0.002)	(0.004)	(0.006)	(0.006)	(0.007)
latitude	0.008	0.025*	0.017	0.049**	0.037*
	(0.005)	(0.013)	(0.020)	(0.020)	(0.021)
MORT_2013	−0.003	0.034**	0.035	0.033	0.028
	(0.006)	(0.014)	(0.022)	(0.022)	(0.023)
Constant	0.241	0.985	2.853**	3.498***	5.338***
	(0.290)	(0.679)	(1.070)	(1.084)	(1.130)
Observations	51	51	51	51	51
R^2	0.076	0.406	0.366	0.536	0.590
Adjusted R^2	0.017	0.368	0.325	0.507	0.564
Residual Std. Error (df = 47)	0.233	0.546	0.861	0.872	0.909
F Statistic (df = 3; 47)	1.286	10.706***	9.029***	18.110***	22.591***

Note: *p<0.1; **p<0.05; ***p<0.01

Trend in Overdose Mortality (2011−2016)





Replication

The replication was partially successful.

Extension

For my extension I define fentanyl exposure differently. Zoorob defines a measure of fentanyl exposure as $Fentanyl_{ij} = \log\left(\frac{S_{ij}}{P_{ij}} + 1\right)$. I define fentanyl exposure several ways:

1. $Fentanyl_{ij} = \log\left(S_{ij} + 1\right)$
2. $Fentanyl_{ij} = \frac{S_{ij}}{\log\left(P_{ij}+1\right)} + 1$
3. $Fentanyl_{ij} = \frac{\log\left(S_{ij}+1\right)}{\log\left(P_{ij}+1\right)}$

Extension 1

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Wed Apr 22 11:38:30 2020

Extension 2

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Wed Apr 22 11:38:31 2020

Extension 3

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Wed Apr 22 11:38:32 2020

These estimates from extensions 1 and 3 are much higher than Zoorob's estimates and the official mortality statistics. Extension 2 is much lower than the official mortality statistics. Zoorob's method of defining fentanyl exposure appears to be the best.

Table 4:

	Dependent variable:				
	ext1				
	(1)	(2)	(3)	(4)	(5)
longitude	0.017**	0.047***	0.053***	0.066***	0.077***
	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
latitude	−0.015	−0.001	−0.049	0.003	−0.009
	(0.027)	(0.030)	(0.036)	(0.035)	(0.035)
MORT_2013	−0.028	0.048	0.029	0.022	0.008
	(0.030)	(0.033)	(0.040)	(0.038)	(0.039)
Constant	5.131***	7.059***	10.730***	10.775***	13.076***
	(1.446)	(1.587)	(1.913)	(1.849)	(1.883)
Observations	51	51	51	51	51
R^2	0.097	0.377	0.359	0.441	0.508
Adjusted R^2	0.040	0.338	0.319	0.405	0.477
Residual Std. Error (df = 47)	1.164	1.277	1.539	1.487	1.515
F Statistic (df = 3; 47)	1.690	9.496***	8.791***	12.348***	16.171***
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 5:

	Dependent variable:	
	age_adjusted_rate	
	(1)	(2)
ext1	2.380**	
	(0.709)	
‘ext1(fit)‘		5.549**
		(1.523)
Observations	357	357
R^2	0.866	0.781
Adjusted R^2	0.840	0.739
Residual Std. Error (df = 299)	7,319.545	9,357.110
Note:	*p<0.1; **p<0.05; ***p<0.01	

	Model 1 Deaths	Model 2 Deaths
2011	19747	46039
2012	20795	48482
2013	23774	55429
2014	31860	74280
2015	39190	91371
2016	46641	108743
2017	52030	121305

Table 6:

	Dependent variable:				
	ext2				
	(1)	(2)	(3)	(4)	(5)
longitude	0.024*	0.241**	0.788**	1.928**	3.422**
	(0.012)	(0.107)	(0.335)	(0.810)	(1.569)
latitude	−0.014	0.149	0.602	1.586	2.200
	(0.039)	(0.340)	(1.064)	(2.569)	(4.977)
MORT_2013	−0.012	0.397	1.119	2.450	5.487
	(0.043)	(0.373)	(1.168)	(2.822)	(5.466)
Constant	4.270**	17.170	53.121	130.373	244.325
	(2.054)	(18.032)	(56.439)	(136.332)	(264.082)
Observations	51	51	51	51	51
R^2	0.079	0.125	0.130	0.130	0.118
Adjusted R^2	0.020	0.069	0.075	0.074	0.061
Residual Std. Error (df = 47)	1.653	14.509	45.411	109.693	212.481
F Statistic (df = 3; 47)	1.342	2.243*	2.342*	2.338*	2.091
Note:	*p<0.1; **p<0.05; ***p<0.01				

I'm still not sure how I'm going to present the uncertainty in these estimates.

Table 7:

	Dependent variable:	
	age_adjusted_rate	
	(1)	(2)
ext2	0.022***	
	(0.004)	
‘ext2(fit)‘		0.060***
		(0.014)
Observations	357	357
R^2	0.902	0.657
Adjusted R^2	0.883	0.592
Residual Std. Error (df = 299)	6,268.497	11,704.250
Note:	*p<0.1; **p<0.05; ***p<0.01	

	Model 1 Deaths	Model 2 Deaths
2011	85	230
2012	86	232
2013	138	374
2014	717	1943
2015	2159	5850
2016	5732	15528
2017	11482	31104

Table 8:

	Dependent variable:				
	ext3				
	(1)	(2)	(3)	(4)	(5)
longitude	0.001*	0.003***	0.003***	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
latitude	−0.0004	0.001	−0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MORT_2013	−0.002	0.004*	0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.300***	0.413***	0.657***	0.640***	0.791***
	(0.088)	(0.092)	(0.116)	(0.105)	(0.102)
Observations	51	51	51	51	51
R^2	0.090	0.433	0.383	0.511	0.600
Adjusted R^2	0.032	0.397	0.344	0.480	0.574
Residual Std. Error (df = 47)	0.071	0.074	0.093	0.084	0.082
F Statistic (df = 3; 47)	1.558	11.982***	9.743***	16.381***	23.501***
Note:	*p<0.1; **p<0.05; ***p<0.01				

Table 9:

	Dependent variable:	
	age_adjusted_rate	
	(1)	(2)
ext3	39.298**	
	(11.594)	
‘ext3(fit)‘		84.072***
		(20.209)
Observations	357	357
R^2	0.870	0.802
Adjusted R^2	0.846	0.765
Residual Std. Error (df = 299)	7,195.770	8,887.805
Note:	*p<0.1; **p<0.05; ***p<0.01	

	Model 1 Deaths	Model 2 Deaths
2011	20136	43079
2012	21218	45393
2013	24225	51827
2014	32521	69575
2015	40023	85623
2016	47610	101854
2017	53079	113554