Milestone 8

Ali Crump

4/24/2020

Abstract

Zoorob (2019) shows that geography and fentanyl exposure explain much of the variation in increased overdose mortality rates between 2011 and 2017. I successfully replicated much of Zoorob's results, but I found discrepancies in the fentanyl exposure coefficients and the total death estimates for each model. My replication finds that the total death estimates are approximately 13% and 16% larger for each model and that the regression coefficients on fentanyl are slightly larger than those published. In addition to replicating Zoorob's work, this paper uses alternate definitions of fentanyl exposure and displays a range of uncertainty in the total estimated deaths while keeping all the other modeling choices the same. The focus of the extension is on Zoorob's ordinary least squares model and I find that the original paper's definition of fentanyl exposure explains more variation in age-adjusted mortality rates than those proposed in the extension. This finding is important because there are many ways to define fentanyl exposure, but Zoorob's appears to be closer.

Introduction

Zoorob (2019) 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:

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

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 (Zoorob (2019)).

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

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. Zoorob 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 (Zoorob (2019)). 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]

For my extension, I use an alternate definition for fentanyl exposure than that proposed by Zoorob: $Fentanyl_{ij} = \log\left(\frac{S_{ij}}{P_{ij}} + 1\right)$. Zoorob does not explicitly explain the decision to define fentanyl exposure using a log transformation other than he aims to create a measure that is comparable across different states and different years Zoorob (2019). He uses population weights in his regression analyses so I propose alternative definitions that do not use the state populations. I also test other definitions that do not use log transformations for all of the variables.

Zoorob's method for measuring fentanyl exposure is the closest to official mortality statistics and explains more variation in mortality rates than the alternate definitions of fentanyl exposure.

Literature Review

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.

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 (Joshua A. Barocas (2020)). Although in 2018, drug overdose deaths actually decreased by 4.6% from 2017 in the United States, fentanyl deaths continued to rise ((*Drug Overdose Deaths*, n.d.))(Abby Goodnough (2019)). (Still looking for literature on this)

Replication

The replication was partially successful. I successfully replicated figures 1, 2, 3, 4, and table 1 from the paper as well figures from the appendix. As an example, Figure 1 is shown in the appendix. However, I was unable to replicate the results from table 2, the parameters of the statistical models of overdose and fentanyl exposure using the ordinary least squares model (Model 1) and the second state of a two-stage least squares model (Model 2). The results for table 2 from the original paper and my replication are shown below:

[1] Github repository

A: Statistical Models of Fentanyl & Overdose		
Dependent variable: Age-Adjusted Mortality Rate		
	(Model 1, OLS)	(Model 2, 2SLS)
Fentanyl Exposure	4.495***	5.426***
	(0.687)	(0.717)
State Fixed Effects	Y	Y
Year Fixed Effects	Y	Y
Population Weights	Y	Y
Observations	357	357
Adjusted R^2	0.915	0.909
First-Stage *F*		31.6

B: Total Estimated Deaths Attributable to Fentanyl by Model.		
	Model 1 Deaths	Model 2 Deaths
2011	2,295	2,705
2012	2,365	2,788
2013	3,312	3,904
2014	8,870	10,458
2015	15,446	18,211
2016	23,188	27,339
2017	30,398	35,841

*p < 0.1.
**p < 0.05.
***p < 0.01.

Zoorob (2019) Table 2 Results.

His model 1 death esimates were 85,874, while my estimates total to 96,969, approximately 13% higher than his. Likewise, for model 2, his estimates total to 101,246, whereas mine total to 117,079, nearly 16% higher. These differenes suggest Zoorob used a different model than the ones included in the replication materials. My estimates for the coefficient on fentanyl exposure are slightly higher than Zoorob's for both models, which might explain why my estimated deaths are also higher. My estimates are shown below.

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sun Apr 26 00:19:04 2020

\begin{center}

Replication results for Table 2. The OLS Model 1 is represented by fentr and the 2SLS Model 2 by fent_r(fit).

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	

	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

\end{center}

Extension

Zoorob defines a measure of fentanyl exposure as $Fentanyl_{ij} = \log\left(\frac{S_{ij}}{P_{ij}} + 1\right)$. He uses this measure of fentanyl exposure in his regression for both models 1 and 2. Zoorob mentions that he uses this measure to compare fentanyl exposure across states of highly variable populations, and that the plus 1 avoids and undefined logarithm of zero. However, he does not provide any further explanation as to why he chose to define fentanyl exposure this way. For my extension, 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. $Fentanyl_{ij} = \frac{\log\left(S_{ij}+1\right)}{\log\left(P_{ij}\right)}$

and run his same analyses. Zoorob’s regression includes population weights and information already, so the first equation challenges the choice of using P_{ij} to define fentanyl exposure. The second definition only takes the log of population, since populations are large compared to the number of fentanyl drug seizures. The third definition is similar to Zoorob’s, but instead takes the log separately for both values, S_{ij} and P_{ij}. The goal of this extension is to see if a different definition of fentanyl exposure is closer to official mortality statistics and is a more reasonable choice.

Table 2:

	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		

Tables

Extension 1

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sun Apr 26 00:19:16 2020

Extension 2

Extension 3

Conclusion

My ordinary least squares estimated fentanyl deaths from extensions 1 and 3, totaling to 234,037 and 238,812 respectively, are much larger than Zoorob's corrected estimates, totaling to 96,969. Zoorob's corrected 2016 estimate, 26,491 deaths, is also more consistent with the official mortality statistic of 19,400 than either of the extension estimates: 46,641, 5,732, and 47,610 respectively. Estimates from extensions 1 and 3 appear to be too high, and estimates from extension 2 appear to be too low to be consistent with official mortality statistics.

Moreover, when comparing R^2 values for each model, Zoorob's method of defining fentanyl exposure appears to explain more of the variation in mortality rate than any of the extensions. Zoorob's R^2 for the OLS model is 0.915, whereas those of the extensions are 0.840, 0.883, and 0.846.

Table 3:

	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 4:

	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 5:

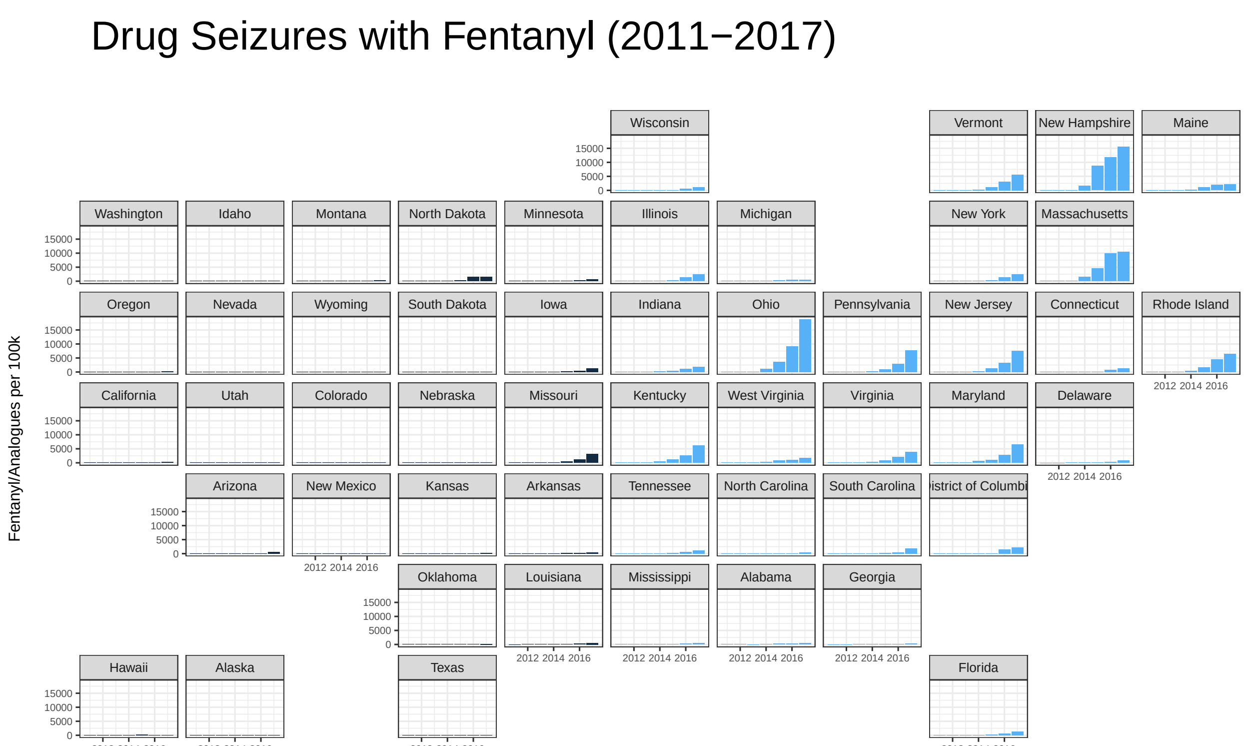
	Dependent variable:				
	ext2				
	(1)	(2)	(3)	(4)	(5)
longitude	0.024*	0.241**	0.788**	1.928**	3.422**
	(0.012)	(0.107)	(0.335)	(0.810)	(1.569)
latitude	−0.014	0.149	0.602	1.586	2.200
	(0.039)	(0.340)	(1.064)	(2.569)	(4.977)
MORT_2013	−0.012	0.397	1.119	2.450	5.487
	(0.043)	(0.373)	(1.168)	(2.822)	(5.466)
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				

Table 6:

	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

Appendix



Source: National Forensic Laboratory Information System (NFLIS)

Table 7:

	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 8:

	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

References

Abby Goodnough, Margot Sanger-Katz, Josh Katz. 2019. *Drug Overdose Deaths Drop in U.s. For First Time Since 1990.* The New York Times. https://www.nytimes.com/interactive/2019/07/17/upshot/drug-overdose-deaths-fall.html.

Drug Overdose Deaths. n.d. Centers for Disease Control; Prevention. https://www.cdc.gov/drugoverdose/data/statedeaths.html.

Joshua A. Barocas, Jianing Wang, Jake R. Morgan. 2020. *Outcomes Associated with Medications for Opiod Use Disorder Among Persons Hospitalized for Infective Endocarditis. Infectious Diseases Society of America.* https://watermark-silverchair-com.ezp-prod1.hul.harvard.edu/ciaa062.pdf?token=
AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAn0wggJ5BgkqhkiG9w0BBwagggJqMIICZgIBADCCA
EQx-LXUgTgLIiPAgEQgIICMC-7gICW1kwD-HanETepgOj5x6deCq5Fr3n9WXriHm-9kI1nEaZorUoyzw9fFR28xBPraP8ad7
x9ke4fdLGmXZ6KXr9G06vH0T45mov4AIT93pZ0u8a1WCqluFe92iBuye2CAUOl0ZOOHmI8fTSPxhwTFSZdu6Uncelk9wrRI
IfiALtxGG5zMzG9jlEnpSaWEmsBeYMmr7IHFRmuKvFkseVcPMtAgtO3N9Wjo7qzWYW48WLME18g2uFRaidSMtzVMCf
DKiIiFGvq7LbF3z0MNRaohlFQ2WpZagYsnscG41hy5OUQJME4K7VpCChcM0V6H159z1wZf4ooVo65ipAjiWs1tgbeXssyjuk
4elwIf0YBWCouCB-vWRbtfJRZnWm4etPfARlgtTKLD4DL4kxW6fes-609BTXekgqTMc0qTqoo4StQUMxGL55SsTayp1EFI
rxb-ASlH8-TZlAKAiAW6RneBSD3ClYw2twDMGAZyEsxNON6veX7tgL3_6WcvJ_zgCUPhBkEkJS22-
03MXt6XY0bOflFasTqjYgRBTAj4pbvHAYpRdRckKWft2iuSPaY06a7zF1P1yah5vWnoPbDvZjlzYC22.

Zoorob, Michael. 2019. *Fentanyl Shock: The Changing Geography of Overdose in the United States. International Journal of Drug Policy.* https://www.sciencedirect.com/science/article/abs/pii/S0955395919301136.