The Labor Supply Consequences of the Opioid Crisis

David Powell

RAND Education and Labor

WR-A1484-1
June 2021

For more information on this publication, visit www.rand.org/pubs/working_papers/WR1484-1.html.

Published by the RAND Corporation, Santa Monica, Calif.

The Labor Supply Consequences of the Opioid Crisis*

David Powell†

First Draft Distributed: June 2020
This Draft: June 2021

Abstract

An emerging literature considers the role of the opioid crisis on labor outcomes, suggesting that increased access to opioids may have led to decreased national labor supply. This paper uses the introduction of OxyContin and geographic variation in its launch to study the labor supply consequences of the opioid crisis. This geographic variation provides an opportunity to study lasting differences in labor supply across states with different exposure to the opioid crisis. This paper uses an event study framework but shows, theoretically and empirically, that a standard event study model with covariates can produce misleading evidence on both the existence of pre-existing trends and post-treatment effects. I implement a simple modification to this standard framework which permits consistent estimation. The results suggest that the opioid crisis played a meaningful role in reducing labor supply for the working-age population.

Keywords: Event Study Designs, Two-way Fixed Effects Models, Labor Supply, Substance Use
JEL Classification: J21, I12, C33

*I thank participants at the University of Pennsylvania Center for Health Incentives & Behavioral Economics, the 2020 APPAM Fall Research Conference, the 2021 American Economics Association Meetings, the 2021 Society of Labor Economists Conference, and the 2021 Econometric Society North American Summer Meetings for helpful discussions and comments. I received especially thoughtful suggestions from Abby Alpert, David Beheshti, Anne Burton, and Atheendar Venkataramani.

†RAND, dpowell@rand.org, (703) 413-1100

1 Introduction

In 2018, more than 67,000 individuals died of drug overdoses, almost 70% of which involved opioids (Wilson, 2020). The opioid crisis is a national emergency, affecting the United States on countless dimensions, and there is widespread interest in understanding its broader effects beyond overdose deaths. In particular, policymakers and researchers are interested in its implications for labor markets. The Federal Reserve Chairman Jerome Powell claimed that the opioid crisis is having a "substantial" effect on the United States economy.[1] Krueger (2017) suggests that the growth in opioid access can explain a meaningful share of the decline in labor force participation.

An emerging literature considers how the opioid crisis has influenced labor supply (Krueger, 2017; Aliprantis et al., 2019; Currie et al., 2019; Savych et al., 2019; Harris et al., 2019; Beheshti, 2019; Park and Powell, 2021; Cho et al., 2021). One motivation for this literature is that the growth of the opioid crisis has occurred as labor force participation rates in the United States have dropped with varying hypothesis offered in the literature (Aaronson et al., 2012; Hotchkiss and Rios-Avila, 2013; Juhn and Potter, 2006; Black et al., 2016; Council of Economic Advisers, 2014).[2] The labor force participation rate and employment-to-population ratio for ages 25-54 are shown in Figure 1 for 1981-2018. Labor force participation peaked in the late 1990's, around the time that overdose rates began to increase rapidly – see Figure A1 for overdose death rates[3] – followed by a relatively steady downward trend in the years since.

The literature has recognized the challenges in isolating growth in opioid access that is unrelated to factors that may independently impact economic conditions. Notably, there is little work studying the relationship between the opioid crisis and national shifts in labor supply back to the origins of the crisis in the 1990s. In this paper, I study changes in labor outcomes from 1981 to 2018 based on state-level variation in initial conditions which exposed some states to the opioid crisis more than others.

[1] `https://www.cnbc.com/2019/07/10/jerome-powell-says-economic-impact-of-opioid-crisis-is-substantial.html`, last accessed April 27, 2020.

[2] Demographic changes, especially population aging, are often found to be important determinants of the fall in labor force participation rates (Fernald et al., 2017; Krueger, 2017). However, labor force participation rates have been found to decline even within younger age groups (Abraham and Kearney, 2018). Abraham and Kearney (2018) find that declines in labor demand, such as those due to import competition and technology, can explain some labor outcome shifts among the working age population.

[3] I provide overdose death rate trends for the full population. The ages 25-54 fatal overdose rates have a similar trend but higher level.

Complementary work shows that the introduction of OxyContin by Purdue Pharma in 1996 explains a large share of the national rise in overdose death rates (Alpert et al., 2021a); my focus on OxyContin is due to its pivotal role in the opioid crisis. OxyContin quickly became a blockbuster drug and Purdue Pharma aggressively marketed the use of strong opioids, such as oxycodone products, more broadly.[4] Within just a few years after its introduction, OxyContin was the most abused opioid (Cicero et al., 2005).

While OxyContin was introduced nationwide, there is persistent geographic variation in opioid supply based on whether a state had a "triplicate prescription program" at the time of OxyContin's launch. These triplicate programs were early and especially stringent forms of prescription drug monitoring programs (PDMPs) which led Purdue Pharma to conclude that "The product should only be positioned to physicians in non-triplicate states..." (Groups Plus, 1995). This differential marketing led to enduring variation in opioid supply, prescribing, promotional activities, and overdose deaths (Alpert et al., 2021a).

A growing literature studies how economic conditions or shocks to labor demand relate to overdoses or a broader set of deaths of despair (Hollingsworth et al., 2017; Ruhm, 2019a; Venkataramani et al., 2020; Pierce and Schott, 2020; Betz and Jones, 2018; Charles et al., 2019; Currie et al., 2019),[5] suggesting that the causal effects of the opioid crisis and economic conditions operate in both directions. More broadly, the "deaths of despair" hypothesis suggests that the opioid crisis is part of a broader trend in deaths related to weakening socioeconomic and cultural conditions for segments of the population (Case and Deaton, 2015, 2017). Other work concludes that the rise in access to opioids (since 2001) has had independent effects on overdose deaths rates (Ruhm, 2019a).

I implement a difference-in-differences design, comparing labor outcomes in triplicate states to outcomes in non-triplicate states both before and after the introduction of OxyContin, relying primarily on event studies to transparently trace the conditional trajectory of these outcomes over time. Given meaningful shifts in demographics between triplicate and non-triplicate states during this time period, I find that it is important to condition on

[4]Purdue Pharma's stated objective in the early years was: "To convince health care professional (physicians, nurses, pharmacists, and managed health care professionals) to aggressively treat both non-cancer pain and cancer pain. The positive use of opioids, and OxyContin Tablets in particular, will be emphasized" (Purdue Pharma, 1999).

[5]Further relevant to this literature, Ruhm (2019b) documents that non-opioid overdose death rates are also rapidly increasing, suggesting that the ongoing epidemic is not specific to opioids. Carpenter et al. (2017) studies prescription pain reliever misuse and economic conditions. Currie and Schwandt (2020) and Maclean et al. (2020) discuss these issues in more detail.

a small set of covariates predictive of labor supply behavior. As shown in Panels C and D of Figure 1, triplicate and non-triplicate states appear to be – unconditionally – on different paths prior to OxyContin's launch which, I find, can be explained by shifting demographics. The traditional approach, implemented regularly in difference-in-differences designs, is to include controls in the specification additively and estimate a homogenous treatment effect.

With event study and difference-in-differences estimates, this approach can induce bias. The covariates partially fit the treatment heterogeneity in the post-period such that changes in covariates over time – even those with no independent relationship with the outcome – potentially lead to bias in event study estimates. I show this bias term theoretically and provide simulation results illustrating the problem.

A recent literature provides a more in-depth understanding of traditional difference-in-differences designs (e.g., Kahn-Lang and Lang (2019); Jaeger et al. (2018)) with special interest in staggered adoption.[6] In this paper, I show that typical implementations of difference-in-differences designs within a regression framework are problematic even without staggered timing.

I also suggest a simple modification to this traditional event study approach. I use only untreated observations to estimate the parameters associated with the covariates as part of a two-step procedure. By selecting only on untreated observations, the relationships between the covariates and outcome are unaffected by treatment effect heterogeneity. After these parameters are estimated, the treatment effects can be estimated using the residualized outcomes.

A methodological contribution of this paper is illustrating how heterogeneous treatment effects can permeate into event study estimates of differences in the *pre-period*, obscuring the true trajectory of differences (or lack of differences) between treated and untreated units prior to treatment.[7] I show that regression adjustments in event study models bias formal and informal (i.e., visual inspection) tests of pre-existing trends since event study estimates will reflect a combination of (1) the true normalized differences between treated and

[6]See de Chaisemartin and D'Haultfoeuille (2020); Goodman-Bacon (2018); Callaway and Sant'Anna (2019); Athey and Imbens (2021); Sun and Abraham (2020); Borusyak and Jaravel (2017).

[7]Recent work re-considers tests of the parallel trends assumptions required for difference-in-differences designs (Roth and Rambachan, 2019; Bilinski and Hatfield, 2020), building on and critiquing a common approach of extending any pre-existing trends linearly into the post-period to predict outcome differences in the absence of treatment.

untreated states and (2) differential shifts in covariates. After using the proposed modified approach, it is straightforward to apply existing techniques to test or relax the parallel trends assumptions. I extend this approach to cases in which all units are partially-treated but vary based on level of treatment (similar to de Chaisemartin and D'Haultfoeuille (2018)).

I also recommend an inference procedure which is suitable for applications with only a few treated units. It relies on a bootstrap-type technique while adjusting for heteroscedasticity differences across units, extending the approach introduced in Ferman and Pinto (2019). In addition, given this paper's focus on covariate adjustment, the inference procedure specifically accounts for the variance of the estimated parameters associated with the covariates.

The analysis relies on labor supply metrics from the Current Population Survey (CPS) with a focus on ages 25-54 (following Krueger (2017)), though I present results for other age groups and data sets as well. Labor supply measures were trending differentially prior to 1996 in triplicate and non-triplicate states, but these trends can be explained by shifting demographics. After adjusting for these demographics, there is no evidence of pre-existing systematic movements prior to 1996 followed by a steady relative increase in labor force participation and employment in triplicate states after treatment through the end of the sample period. The estimates imply that triplicate states, due to their reduced exposure to OxyContin's launch, experienced relative labor force participation growth of about 2 percentage points and employment-to-population growth over 3 percentage points for the 1996-2018 period. These magnitudes are large but imply relationships between opioid access and labor supply outcomes at the lower end of the spectrum of estimates found in the literature.

I observe similar differential growth for other labor supply measures such as hours worked and earnings, labor supply for broader age groups, and establishment-level measures of employment. The results cannot be explained by systematic labor demand shocks, other policies targeting opioid access and misuse, functional form assumptions, or migration. I also provide estimates using a traditional difference-in-differences approach. This traditional approach produces different results and we would likely reach different conclusions using this more restrictive design.

I provide more background about the literature's findings on the relationship between opioid access and labor supply in the next section as well as additional information on OxyContin and Purdue Pharma's launch plan. Section 3 evaluates how traditional two-

way fixed effects models operate when covariates are added while also discussing a modified approach and an inference procedure. Section 4 discusses the empirical strategy and data. Section 5 presents the results and Section 6 concludes.

2 Background

2.1 Labor Supply Effects of Opioid Access

Much of the literature on the impact of opioid prescribing (or over-prescribing) on labor supply relies on geographic-specific changes or cross-sectional variation, assuming that such variation occurs for reasons unrelated to labor market conditions.[8] It is rare in this literature to leverage policy-driven variation.[9] A recent exception is Beheshti (2019), which exploits the differential geographic impacts of the rescheduling of hydrocodone, finding that reduced medical access to hydrocodone due to rescheduling improved labor force participation rates. Similarly, Park and Powell (2021) study the reformulation of OxyContin, the removal of one of the most highly-abused opioids, finding sizable reductions in labor supply as workers shifted to illicit drug markets. Cho et al. (2021) observe similar consequences of the reformulation of OxyContin.

Shifts in opioid access in more recent years may have different consequences than variation at the beginnings of the opioid crisis. For example, Beheshti (2019) and Park and Powell (2021) consider the consequences of reducing access to a substance or a highly-abusable version of a substance at a time when illicit opioids were readily-available. The launch of OxyContin would potentially have different effects since illegal opioid markets had yet to develop and misuse of opioids was not as widespread. The lack of quasi-experimental evidence on the long-run labor supply effects of the opioid crisis is due to the inherent difficulties in finding variation in exposure to the opioid crisis since its origins.

The literature generally tries to understand the net effect of increased access to opioids to quantify the impact of the opioid crisis on labor markets. This paper is motivated by the same goal. I study a large differential shock to opioid access and evaluate how changes in opioid supply lead to net changes in labor outcomes. Increased access to opioids,

[8] Alpert et al. (2021b) is able to study measures of labor productivity at the individual-level more directly using matched military labor and health records and the random assignment of emergency department physicians. Thingholm (2020) uses physician-level variation in Danish prescribing habits to study effects on labor productivity. Laird and Nielsen (2016) use a mover strategy to identify the labor supply effects of opioid prescribing rates on Danish labor outcomes.

[9] A small, related literature considers the impact of prescription drug monitoring programs on labor supply outcomes (Kaestner and Ziedan, 2019).

in principle, has an ambiguous impact on labor supply behavior. Opioid use may lead to dependence issues, reducing productivity and the ability to work. Alternatively, opioid access may improve pain management and increase labor force participation rates.[10] There are also potential general equilibrium effects involved – for example, the rise in opioid dependence may strain health care systems, expand illicit markets and increase crime, and have broader impacts on the local economy. Thus, labor supply effects are not necessarily directly related to individual misuse but may involve sizable spillovers. Magnitudes in the literature suggest sizable general equilibrium effects.[11]

2.2 OxyContin's Launch

To explore the effects of the opioid crisis on labor outcomes, I focus on the role of OxyContin. Recent work concludes that 80% of the rise in the overdose death rate since 1996 can be attributed to the introduction and marketing of OxyContin (Alpert et al., 2021a). OxyContin was approved by the Food and Drug Administration (FDA) in 1995 and introduced to the market in January 1996 by Purdue Pharma. The key innovation of OxyContin was its long-acting formula which provided 12 hours of continuous pain relief, significantly improving the quality and ease of pain management over previous drugs. However, the timed-release aspect of OxyContin is contingent on taking the pill whole. Crushing or dissolving the pill causes the high dose of oxycodone, intended to be released slowly over 12 hours, to be delivered all at once. This property made OxyContin especially easy to abuse.

2.2.1 OxyContin Promotion

To understand the initial marketing of OxyContin, I made Freedom of Information Act (FOIA) requests to Florida, Washington, and West Virginia to obtain recently unsealed documents from investigations and court cases brought against Purdue Pharma in these states. These documents provide a rare look at a pharmaceutical firm's detailed marketing strategies. An example can be observed in Figure A2.

These internal Purdue Pharma documents included survey research suggesting that triplicate prescription programs had a chilling effect on the prescribing of strong opioids because physicians were worried about government oversight and due to the additional hassle

[10] Currie et al. (2019) find positive labor supply effects for younger women

[11] For example, the effect sizes in Park and Powell (2021) are larger than the OxyContin misuse rate. This disproportionate effect may suggest that the overall OxyContin misuse rate – the predictor of exposure to reformulation in Park and Powell (2021) – in the National Survey on Drug Use and Health is severely under-reported, a possibility that I cannot rule out. Alternatively, the results suggest that the general equilibrium effects of large increases in substance use are sizable.

of the triplicate forms.[12]

Purdue Pharma's launch plan mentions triplicate programs dozens of times, acknowledging that "these regulations create a barrier when positioning OxyContin" (Purdue Pharma, 1995). Since there would be lower returns to promoting OxyContin in triplicate states, they recommended that "the product [OxyContin] should only be positioned to physicians in non-triplicate states" (Groups Plus, 1995).

2.2.2 Triplicate Prescription Programs

Triplicate prescription programs emerged as some of the earliest prescription drug monitoring programs to monitor the prescribing and diversion of controlled substances. In a triplicate prescription program, the prescriber was mandated to use state-issued triplicate prescription forms when prescribing Schedule II controlled substances. The prescriber keeps one copy of the prescription while the patient provides the remaining two to the pharmacy, which keeps one copy while sending the other to the state monitoring agency.

At the time of OxyContin's launch, five states had triplicate programs – California (enacted 1939), Idaho (1967), Illinois (1961), New York (1972), and Texas (1982). Interestingly, these programs were enacted decades prior to the beginnings of the opioid crisis and were phased out in the years following OxyContin's launch. Thus, this study examines the longer-term consequences of the initial targeting of OxyContin.[13]

Figure A3 provides evidence that non-triplicate states were more exposed to the introduction of OxyContin. In Panel A, I study promotional payments for OxyContin, obtained from the Open Payments data base for August 2013 to the end of 2016. There are large differences in promotional activities for OxyContin between triplicate and non-triplicate states. It might be surprising that prescribing regulations in 1996 could have such persistent effects on marketing strategies. A defining feature of Purdue Pharma's detailing strategy was its persistence. Internal documents suggest that a primary strategy of the sales force was to call and visit the top OxyContin prescribers, and this behavior continued (and

[12] "Writing triplicate prescriptions was more trouble than others, due to the details of the forms and the various people that need to be copied to them. To the extent that they [physicians] can avoid this extra effort, they will try to follow alternative protocols" (Groups Plus, 1995).

[13] Alpert et al. (2021a) dedicates a significant amount of analysis to isolating the underlying mechanisms explaining the large differential post-1996 overdose trends, determining that it is primarily driven by a marketing effect and not the direct effects of triplicate programs themselves. In this paper, I use the 1996 launch of OxyContin as a large, differential shock to opioid access to study the labor supply consequences. The exact mechanism (marketing versus triplicate programs encouraging persistent prescribing habits) are of less interest in this context.

even became more frequent) until recently.[14] Thus, contemporary differences in promotional activities reflect variation in initial targeting interacted with a marketing strategy promoting targeting of high prescribers (who were likely previously targeted).

I also show early and longer-term differences in OxyContin "adoption" as measured in terms of prescriptions and morphine equivalent doses. I provide differences in per capita OxyContin supply (morphine equivalent doses) from the ARCOS data for 2000-2016 (Panel B),[15][16] per beneficiary Medicaid prescriptions from the SDUD data for 1997-2005 (Panel C),[17] and per capita prescriptions in the MEPS for 1996-2016 (Panel D).[18] In all cases, I observe triplicate states with substantially less access/use of OxyContin, both in the years after launch and persisting through the most recent years of data.

3 Difference-in-Differences with Covariates

The section considers the common practice of including covariates linearly in a two-way fixed effects specification in the presence of treatment effect heterogeneity. The motivation for this analysis is due to the frequency that this approach is used in applied work. When discussing covariates in the context of difference-in-differences, Cameron and Trivedi (2005) state, "The standard solution is to include such controlling variables in the regression" (page 57).

Some work has suggested semi-parametric estimators, flexibly estimating the relationship between covariates and the outcome separately for each interaction of treated/untreated and pre/post (Blundell et al., 2004; Abadie, 2005). Sant'Anna and Zhao (2020) discuss that traditional difference-in-differences assumes no covariate-specific trends in the treatment units relative to the control units. In addition, de Chaisemartin and D'Haultfoeuille (2020) includes a covariate adjustment similar to the approach suggested here. In this paper,

[14]Later years saw increased targeting to high prescribers. For example, "McKinsey recommended doubling down on Purdue Pharma's strategy of targeting high prescribers for even more sales calls..." (Commonwealth of Massachusetts, 2018).

[15]The online ARCOS data are provided by ingredient. I made a FOIA request for data on OxyContin specifically. This request was approved for 2000-2016. I was told that earlier data were unavailable.

[16]There is a notable dip in national OxyContin supply in 2005-2006 due to a patent dispute at this time. Alpert et al. (2021a) show that total oxycodone supply was unaffected over this period, presumably as generic OxyContin formulations substituted for the brand-name version. This dispute was eventually resolved and the generic versions exited the market. Despite these large shifts in national trends, the relative supply of non-triplicates and triplicates remained approximately the same.

[17]Since Medicare Part D was enacted in 2006, I end the time series in 2005. While the current SDUD available online suppresses numbers of prescriptions less than ten, I use a version downloaded before this suppression policy was implemented.

[18]I accessed these data in the AHRQ Data Facility due to the necessity for geocoded data.

I pay special attention to the implications of additive covariates for event studies and the evaluation of pre-trends and post-treatment dynamic effects.[19]

I begin with a simple model for illustrative purposes. I assume that all units adopt at the same time and there are T_0 periods prior to treatment. For notation, define W_s as an indicator equal to 1 if unit s is ever treated, such that the treatment variable is $D_{st} = W_s \times \mathbf{1}(t > T_0)$. For now, assume the true model is

$$\text{True Model:} \quad y_{st} = \alpha_{s0} + \gamma_{t0} + \beta_0 D_{st} + \delta_0 X_{st} + \phi_0 D_{st} X_{st} + \epsilon_{st}, \tag{1}$$

where y_{st} represents the outcome for unit s at time t. A single covariate, X_{st}, has an independent effect on the outcome but also interacts with the treatment variable. I index all parameters above with "0" to signify that these are the true parameters. I evaluate what happens when the estimated specification does not include the interaction term. I assume an event study model is estimated:

$$\text{Estimated Event Study Specification:} \quad y_{st} = \alpha_s + \gamma_t + \beta_t W_s + \delta X_{st} + \varepsilon_{st}. \tag{2}$$

The estimated model does not include interactions between D and X, which will lead to omitted variable bias. Given equation (1), we want estimates such that $E[\widehat{\beta}_t | t \leq T_0] = 0$, $E[\widehat{\beta}_t | t > T_0] = \beta_0 + \phi_0 E[X_{st} | D_{st} = 1]$.

However, we estimate something different with equation (2). To evaluate the bias, first consider a regression of DX on X, D, unit fixed effects, and time fixed effects; define μ as the parameter on X in this regression. Technically, μ can be zero (or even negative), but we generally expect $\mu > 0$ given the mechanical relationship between DX and X. The omitted variable bias formula tells us that $E[\widehat{\delta}] = \delta_0 + \phi_0 \mu$. This bias creates problems because the event study estimates incorrectly reflect changes in $y - \widehat{\delta} X$, which is not the same in expectation as $y - \delta_0 X$ when $\phi_0 \mu \neq 0$. Let us evaluate outcome differences between treated and untreated units and how these differences evolve over time (for some t_0 and t_1) according to the *estimated* event study model in equation (2):

$$\Big(E[y_{st} | W_s = 1, t = t_1] - E[y_{st} | W_s = 0, t = t_1] \Big) - \Big(E[y_{st} | W_s = 1, t = t_0] - E[y_{st} | W_s = 0, t = t_0] \Big) \tag{3}$$

[19] A large literature discusses this issue and "full regression adjustment" in the context of experiment data. See Negi and Wooldridge (2020) for a recent paper.

$$= (\beta_{t_1} - \beta_{t_0}) + \underbrace{\phi_0 \mu \Big\{ \Big(E[X_{st}|W_s = 1, t_1] - E[X_{st}|W_s = 0, t_1] \Big) - \Big(E[X_{st}|W_s = 1, t_0] - E[X_{st}|W_s = 0, t_0] \Big) \Big\}}_{\text{bias term}}.$$

δ is asked to fit the relationship between y and X for both treated and untreated units. It cannot accomplish both, inducing bias. The bias is equal to the size of the heterogeneous treatment effects (ϕ) times the conditional relationship between DX and D, multiplied by the differential trend in the covariate.

3.1 Simulations

To further illustrate these problems, I simulate data for $T = 30, N = 50$. I consider the last 10 periods treated (i.e., $T_0 = 20$) for the first 5 units: $d_{st} = \mathbf{1}(t > 20, s \leq 5)$. In addition, there is a covariate correlated with treatment. Define $e_{st} \sim N(0, 1)$ and the covariate as:

$$x_{st} = \left\{ \begin{array}{lcc} e_{st}, & \text{for } s > 5 & \text{(Untreated Units)} \\ t + e_{st}, & \text{for } t \leq 20, s \leq 5 & \text{(All Treated Units, Pre-Period)} \\ -1, & \text{for } t > 20, 1 \leq s \leq 2 & \text{(2 Treated Units, Post-Period)} \\ 0, & \text{for } t > 20, s = 3 & \text{(1 Treated Unit, Post-Period)} \\ 1, & \text{for } t > 20, 4 \leq s \leq 5 & \text{(2 Treated Units, Post-Period)} \end{array} \right\}$$

The outcome is $y_{st} = s + t + d_{st}x_{st} + \varphi_{st}$, where $\varphi_{st} = 0.2\varphi_{s,t-1} + 0.5\eta_{st}$ with $\eta_{st} \sim N(0, 1)$. The average treatment effect on the treated (ATT) is equal to zero in each treated period since $E[x_{st}|d_{st} = 1] = 0$. I simulate these data 10,000 times and estimate a traditional event study, presenting the mean time-specific estimates graphically. Since the true effect in each period is equal to zero, these averages can be considered bias estimates.

I show the results in Figure 2. Panel A provides the mean estimates for a traditional event study (normalized to zero in the period prior to treatment). The event study estimates suggest that there is a steep downward time trend before treatment. At the time of treatment, we observe a dramatic increase. Visual inspection of these estimates suggest that the policy had a large effect. Researchers in such cases may even consider adjusting for the pre-trend in some manner, likely increasing the policy estimate even further.

However, neither the pre-trend nor the "policy effect" beginning at time $T_0 + 1$ is real. Instead, both solely reflect trends and shocks to the *covariate* due to interactions of

the treatment and the covariate.[20] This example is intentionally extreme to illustrate the underlying issue, but equation (3) shows that such bias can occur more generally.

3.2 Modified Approach

Define $\hat{\alpha}_s, \hat{\gamma}_t$, and $\hat{\delta}$ as the estimated parameters resulting for a regression of y on state fixed effects, time fixed effects, and X selected only on untreated observations (i.e., $D_{st} = 0$). The idea is to use the untreated observations to estimate the relationships between the covariates and the outcome. Define $\hat{\theta}_{st} \equiv y_{st} - \hat{\alpha}_s + \hat{\gamma}_t + \hat{\delta} X_{st}$. Under typical assumptions, such as exogeneity of treatment, discussed more formally later, then for equation (1),

$$E[\hat{\theta}_{st}|W_s = 1, t \leq T_0] - E[\hat{\theta}_{st}|W_s = 0, t \leq T_0] = 0$$
$$E[\hat{\theta}_{st}|W_s = 1, t > T_0] - E[\hat{\theta}_{st}|W_s = 0, t > T_0] = \beta_0 + \phi_0 E[X_{st}|D_{st} = 1]$$

The first equation represents the estimate for β_t when $t \leq T_0$. These are the pre-period event study estimates. The second equation is the estimate when $t > T_0$. $\beta_0 + \phi_0 E[X_{st}|D_{st} = 1]$ is the ATT. By regressing y on covariates using untreated observations only, those estimates are not impacted by any interactions with D since $D = 0$.

The rest of this section defines the parameters of interest with special attention to the case (relevant to this paper) of unit*time data with different population sizes. Let $\mathcal{S}_1$ represent the set of treated units; $\mathcal{S}_0$ represents the set of untreated units. Define M_{jt} as the population size of unit j at time t and further define $W_{Post,Treated} \equiv \sum_{t=T_0+1}^{T} \sum_{s \in \mathcal{S}_1} M_{st}; W_{Pre,Treated} \equiv \sum_{t=1}^{T_0} \sum_{s \in \mathcal{S}_1} M_{st}; W_{Post,Control} \equiv \sum_{t=T_0+1}^{T} \sum_{s \in \mathcal{S}_0} M_{st}; W_{Pre,Control} \equiv \sum_{t=1}^{T_0} \sum_{s \in \mathcal{S}_0} M_{st}$.

The difference-in-differences estimate is

$$\hat{\beta} \equiv \sum_{j \in \mathcal{S}_1} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Post,Treated}} - \sum_{t=1}^{T_0} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Pre,Treated}} \right) - \sum_{j \in \mathcal{S}_0} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Post,Control}} - \sum_{t=1}^{T_0} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Pre,Control}} \right). \tag{4}$$

Alternative definitions are possible since other weighting schemes may be more desirable. For example, it is possible to population-weight *within* a year but then equally-weight each of the year-specific estimates. In the application of this paper, these alternative estimators produce

[20] In this case, not controlling for x – since it does not have an independent effect on y – produces unbiased estimates (simulation results are provided in Figure A4). Of course, in cases in which the covariates independently influence the outcome (and are correlated with treatment), then excluding covariates will also lead to bias.

nearly-identical results. I define the event study year-specific estimates comparably:

$$\hat{\beta}_t \equiv \sum_{j \in \mathcal{S}_1} \left(\frac{M_{jt}\hat{\theta}_{jt}}{W_{Post,Treated}} - \frac{M_{jt}\hat{\theta}_{jt}}{W_{Pre,Treated}} \right) - \sum_{j \in \mathcal{S}_0} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Post,Control}} - \sum_{t=1}^{T_0} \frac{M_{jt}\hat{\theta}_{jt}}{W_{Pre,Control}} \right). \quad (5)$$

In a traditional event study, the estimates are relative to an omitted period, such as the time period prior to adoption. I have defined the baseline as the entire pre-period (similar to difference-in-differences estimates). The estimation steps are as follows:

1. Selecting only on observations in which $D_{st} = 0$, regress y on unit fixed effects, time fixed effects, and $\boldsymbol{X}_{st}$.
2. Residualize every observation in the data using the estimates from step 1.
3. Estimate equation (4) and equation (5) for all t.

The formal assumptions underlying this approach are discussed in Appendix Section B.1. The problems discussed at the beginning of this section are caused by the interaction of X and D and not explicitly including this interaction in the specification. In the simple example discussed in Section 3, including that interaction would solve these problems. However, explicitly including these interaction terms may be prohibitive as the model gets more complicated. For example, there may be many covariates but few treated units. Additionally, treatment effects may also vary over time, further complicating explicit estimation of this heterogeneity. The residualization approach appropriately handles these other cases.

3.3 Simulation Results using Modified Approach

Adopting this modified approach, I revisit the simulations from Section 3.1. I show the results in Figure 2, Panel B. I observe estimates hovering around zero throughout the entire time period. There is no evidence of a pre-existing trend or post-treatment bias. By selecting only on untreated observations, the model is able to account for the independent effects of the covariate in a manner unaffected by treatment heterogeneity.

3.4 Residualization

A common trends assumption is required for the residualized outcomes (see Appendix Section B.1) and relies on obtaining an estimate of $E[\boldsymbol{\gamma}|W_s, t]$, the relationship between the covariates and the outcome. It is typical in applied work to assume that the covariates' relationships with the outcome are constant across units and time. However, there are gains in permitting this relationship to vary by treatment status and time, such that $E[\boldsymbol{\gamma}|W_s, t] = \boldsymbol{\gamma}_{wt}$. This flexibility is especially important for empirical applications

with differential degrees of treatment (see Section 3.6 below). It is straightforward to permit treatment-specific heterogeneity since treated and untreated units are observed prior to treatment. Time-specific heterogeneity (for the post-period) is also possible to estimate using the untreated observations in post-treatment period.[21] We can assume that there is an additive time-specific component to the expected value of $\boldsymbol{\gamma}$ such that $E[\boldsymbol{\gamma}|W_s, t] = \boldsymbol{\gamma}_w + \boldsymbol{\zeta}_t$.[22] The idea is that the untreated units in the post-period identify *changes* to the returns to the covariates with an assumption that, in the absence of treatment, the treated units would have experienced the same changes.[23]

The assumption of this approach is that the treated and untreated units did not experience differential shocks to the returns to any of the covariates post-treatment: $E[\boldsymbol{\zeta}_t|W_s = 1, t > T_0] = E[\boldsymbol{\zeta}_t|W_s = 0, t > T_0]$, a standard exogenity assumption. One concern with this approach is that it may require the inclusion of a large number of interactions approaching the number of untreated units. In such cases, it may be appropriate to use regularlization/penalization techniques to reduce concerns of over-fitting.

3.5 Inference

Given a small number of treated units and possible heterogeneous treatment effects, inference in this context can be difficult. For example, some inference procedures, such as the wild bootstrap, assume homogeneity of the design matrix (Canay et al., 2019), which rules out difference-in-differences designs. Other inference methods are tailored for difference-in-differences designs but assume homogeneous treatment effects (Canay et al., 2017; Ibragimov and Müller, 2016). I use an inference procedure related to the method introduced in Ferman and Pinto (2019) but extended to the proposed estimation approach.

I focus on inference of β, but the results in this section also apply to β_t. Consider the null hypothesis $H_0 : \beta = \beta_0$. For the base case, I assume that the unit-time data were aggregated from individual-level weighted data with weights ω_{ist} for N_{st} observed observations such that $M_{st} \equiv \sum_{i=1}^{N_{st}} \omega_{ist}$.

I randomize, with replacement, untreated units into placebo "treated" and "un-

[21] Jaeger et al. (2018) suggest interacting pre-treatment covariates with unit-specific trends or time dummies. It is not necessary to fix the covariates to pre-treatment values, though there may be benefits to this depending on the context.

[22] As one simplification, it is reasonable to restrict $\boldsymbol{\zeta}_t = 0$ for $t \leq T_0$ given that we are less worried about heterogeneity within the untreated period.

[23] Alternatively, it is possible to "forecast" this relationship by, for example, permitting the relationship between the covariates and outcome for the untreated units to follow a linear trend, estimating that trend in the pre-period, and then projecting to the post-period.

treated" units to estimate $\hat{\beta}_b$. Each unit j is mapped to $b(j)$. I do *not* use the treated units to construct placebo estimates. This constraint is necessary because the null hypothesis relates to the average (across units) effect and does not provide information about unit-specific effects. If we split the treated units into placebo treated units and placebo control units, then the placebo control units – even after imposing the null – may have a non-zero impact.[24] The placebo estimate is:

$$\hat{\beta}_b \equiv \sum_{j \in \mathcal{S}_1} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\hat{\theta}_{b(j)t}}{W_{Post,Treated}} - \sum_{t=1}^{T_0} \frac{M_{jt}\hat{\theta}_{b(j)t}}{W_{Pre,Treated}} \right) - \sum_{j \in \mathcal{S}_0} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\hat{\theta}_{b(j)t}}{W_{Post,Control}} - \sum_{t=1}^{T_0} \frac{M_{jt}\hat{\theta}_{b(j)t}}{W_{Pre,Control}} \right). \tag{6}$$

The unit*time-specific weights are held constant and $\hat{\theta}_{jt}$ is replaced by $\hat{\theta}_{b(j)t}$. This property is important for permitting estimation of the variance. The variances of $\hat{\beta}$ and $\hat{\beta}_b$ are not necessarily equal given that the treated and untreated groups may have different population sizes.

Note that $\hat{\theta}_{st} = \epsilon_{st} + (\alpha_s - \hat{\alpha}_s) + (\gamma_t - \hat{\gamma}_t) + \boldsymbol{X}'_{st}\left(\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}\right)$. Define the residual in the following manner: $\epsilon_{st} = \nu_{st} + \sum_{i=1}^{N_{st}} \frac{\omega_{ist}}{M_{st}} \eta_{ist}$. There is a unit-year component and an individual component. I assume that the ν_{st} terms are serially-correlated over time with $\nu_s \equiv (\nu_{s1}, \ldots, \nu_{sT})$ i.i.d across s. The η_{ist} terms are i.i.d. Define

$$\Delta \boldsymbol{X}_b \equiv \sum_{j \in \mathcal{S}_1} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\boldsymbol{X}_{b(j)t}}{W_{Post,Treated}} - \sum_{t=1}^{T_0} \frac{M_{jt}\boldsymbol{X}_{b(j)t}}{W_{Pre,Treated}} \right) - \sum_{j \in \mathcal{S}_0} \left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\boldsymbol{X}_{b(j)t}}{W_{Post,Control}} - \sum_{t=1}^{T_0} \frac{M_{jt}\boldsymbol{X}_{b(j)t}}{W_{Pre,Control}} \right).$$

The variance can be written as $\text{Var}(\hat{\beta}_b) \equiv A + Bq_b + \Delta \boldsymbol{X}_b \Delta \boldsymbol{X}'_b \text{Var}(\hat{\boldsymbol{\delta}})$. The above equation defines $\text{Var}(\hat{\beta})$ given $b(j) = j$. This result, and the definition of q_b, is discussed in detail in Section B.2. q_b is a function of population size and weights and is the factor which adjusts for heteroscedasticity. I estimate the parameters A and B in a regression of $\hat{\beta}_b^2$ (or $(\hat{\beta} - \beta_0)^2$) on a constant and q_b. $\Delta \boldsymbol{X}_b \Delta \boldsymbol{X}'_b \text{Var}(\hat{\boldsymbol{\delta}})$ is also included on the right hand side, but its effect is constrained to equal one.[25] I use the traditional cluster covariance matrix estimator (i.e., "cluster by state") to estimate the variance ($\widehat{\text{Var}(\hat{\boldsymbol{\delta}})}$), assuming that $\hat{\boldsymbol{\delta}}$ is estimated using a large number of comparison units.

[24] Alternatively, one could select all the treated units jointly (with replacement) to be part of the control group (or as the treated group).

[25] Variance estimates related to the estimation of the unit and time fixed effects are ignored because these terms cancel out in equation (4).

We expect both A and B to be positive. I constrain these parameters to be non-negative to avoid the possibility of estimating negative variances. In practice, these constraints do not impact the findings in the paper.[26]

Estimating the variance of the aggregate estimate, instead of state-specific estimates, has the advantage that it is straightforward to adjust for the variance of associated with $\hat{\boldsymbol{\delta}}$). State-specific estimates may be correlated (positively or negatively) with one another in finite samples due to values of $\boldsymbol{X}$. The construction of $\Delta\boldsymbol{X}_b$ accounts for these correlations.

After estimation of the parameters A and B, I construct estimates of the variance for each $\hat{\beta}_b$, represented by $\widehat{\text{Var}(\beta_b)}$. Then, I define $\tilde{\beta}_b = \hat{\beta}_b \frac{1}{\sqrt{\widehat{\text{Var}(\beta_b)}}}$. The value of $\hat{\beta}\frac{1}{\sqrt{\widehat{\text{Var}(\beta)}}}$ within the placebo distribution defines a p-value. The steps are as follows:

1. Calculate DID estimate $\hat{\beta}$.
2. Create $\mathcal{B}$ samples and estimates using equation (6). For each iteration:
   (a) Resample with replacement N times from untreated units only
   (b) Assign (with replacement) $|\mathcal{S}_1|$ units as "treated"
   (c) Calculate estimate $\hat{\beta}_b$
3. Using all placebo samples, estimate A and B.
4. Rescale estimates by predicted variance: $\frac{\hat{\beta}_b}{\sqrt{\widehat{\text{Var}(\hat{\beta}_b)}}}$
5. Reject H_0 at level α if and only if $\frac{|\hat{\beta}|}{\sqrt{\widehat{\text{Var}(\hat{\beta})}}} > \frac{|\hat{\beta}_b|}{\sqrt{\widehat{\text{Var}(\hat{\beta}_b)}}}[1-\alpha]$

This approach shares many of the properties of the Ferman and Pinto (2019) method. The primary difference is that Ferman and Pinto (2019) use unit-specific differences (post minus pre), rescale those to adjust for heteroscedasticity, and then aggregate to construct placebo estimates. Here, I propose constructing the placebo estimates and rescaling using the predicted variance of this aggregated estimate. One advantage of this order of implementation is that it is straightforward to adjust for the variance of the estimated parameters associated with that covariates.[27]

[26] Ferman and Pinto (2019) propose a finite sample adjustment when one or both of these parameters are negative. Confidence intervals throughout this paper are similar if this finite sample adjustment is followed.

[27] Ferman and Pinto (2019) point out that this term asymptotes to zero, though I find it is an important adjustment in finite samples.

3.6 Differential Levels of Treatment

I have focused so far on difference-in-differences applications in which some units are untreated at all times. There are many applications in which all units are initially untreated and then later treated with varying intensities (e.g., Alpert et al. (2018); Kearney and Levine (2015); Andersen et al. (2020)). The proposed approach works in such cases as well, and the resulting estimates represent the causal change between more treated and less treated units.

To recover the "average treatment effect on the more treated" it is necessary to construct an appropriate counterfactual for the more treated units. To the extent that there are heterogeneous effects dependent on covariates, then permitting time-specific covariate effects among the "less treated" group accounts for the impacts of treatment for these units as well as any secular changes in the independent effects of the covariates. This flexibility permits estimation of the counterfactual for the more treated units if they had been less treated. The assumption, however, is that all sources of heterogeneity are accounted for by including these covariate*time interactions. This assumption is difficult to test. For most of this paper, I interpret the results as estimates of the causal changes between more and less treated units. Appendix Section B.3 discusses this issue further as well and provides simulation results for this type of setup.

4 Data and Empirical Approach

4.1 Data

The primary data set of this paper is the Current Population Study (CPS) for 1981-2018 (Flood et al., 2020). The CPS has several advantages in this context. First, the CPS provides relatively consistent measures of labor supply over a long time period. Second, it provides information on different dimensions of labor supply including labor force participation and hours worked. Finally, it provides demographic information which allows me to construct the measures of interest for ages 25-54 as well as other age groups. It also provides information on sex, race, ethnicity, and education. These demographic variables are helpful for predicting changes in labor supply.

The variables of interest are included in both the Basic Monthly files and the Annual Social and Economic Supplement (ASEC). To improve power, I use the ASEC data and the monthly files together, and treat the ASEC sample as a separate month (i.e., I estimate a separate time effect for each ASEC sample year). Since I provide annual event study

estimates, this approach does not create any interpretation difficulties. I will also provide estimates without the ASEC and for an ASEC-only sample.

I focus primarily on labor force participation, defined as people with a job or looking for a job in the preceding week. I also separately study the share working in the previous week.[28] As a complementary metric, I study hours worked in the previous week across all jobs. This variable is topcoded to 99 until January 1994 when the topcode is increased in non-ASEC samples. I topcoded these later data to 99 for consistency purposes, affecting 0.03% of individual-level observations.

I also study annual labor earnings, defined as the sum of wages, business income, and farm income. These individual components are each censored at high values and the censoring points change throughout our time period. I do not adjust for censoring, but I will study a complementary outcome on employee compensation using Bureau of Economic Analysis (BEA) data which is not affected by censoring. Because earnings refer to the previous calendar year, I use the 1982-2019 CPS samples to construct earnings measures for 1981-2018. Earnings are expressed in 2019 dollars.

The education variable in the CPS changes in 1992. Prior to 1992, there is less information on degree attainment. I construct education measures in a manner to maximize comparability by focusing on number of years completed. My analysis uses "no college education" and "at least some college education," though I provide results using a richer characterization as well.[29] I observe small discontinuities in 1992 in the constructed measure, but they do not appear to impact the main event studies.

For the primary analysis, I follow Krueger (2017) and select on the 25-54 age group, though I also show results for other age groups.[30] I use the CPS weights to construct means by gender, state, and time.

[28]There was a survey redesign in 1994 which potentially affected responses to the employment questions (Polivka and Miller, 1998). I assume that time fixed effects account for such changes. I do not observe any notable discontinuities between triplicate and non-triplicate states in any labor outcome measures in 1994 so I conclude that these time effects are sufficient.

[29]Less than 12, 12+ but no college, some college but less than 5 years, and 5+ years of post-secondary education.

[30]I exclude respondents listed as in the Armed Forces.

4.2 Empirical Approach

I follow the approach discussed in Section 3.2 and model untreated outcomes as

$$Y_{sgmy}(0) = \alpha_{sg} + \delta_{gmy} + \boldsymbol{X}'_{sgmy}\boldsymbol{\gamma}_{gw(s)y} + \epsilon_{sgmy},$$

where $Y_{sgmy}(0)$ represents the "untreated" labor outcome for state s in month m and year y for gender g. All parameters vary by gender since we suspect that the dynamics of labor supply are different for men and women. The specification includes state-gender and time (month-year)-gender fixed effects. Each observation is a state*gender*time cell.

I consider states prior to 1996 as untreated and triplicate states as "treated" for 1996-2018. Thus, the treatment is *less* exposure to OxyContin's launch with non-triplicate states acting as the more exposed counterfactual. In principle, it does not matter which group is considered treated or untreated. In practice, denoting the triplicate states as treated is beneficial since the parameters above are estimated using the non-treated states and it is helpful to have a large set of control units.

I provide results in which $\boldsymbol{\gamma}_{gw(s)y}$ varies by gender-triplicate status (where $w(s)$ denotes the triplicate status of state s). I also permit the parameters to vary by gender-year in the treated period, using the non-triplicate states to estimate how these parameters evolve over time.[31] Because I permit $\boldsymbol{\gamma}_{gw(s)y}$ to vary over time, it is important to avoid including a large set of covariates in X. I select a small set of covariates, discussed in the next section, which are likely correlated with labor supply outcomes and because they have differential trends by treatment status. However, I also explore the consequences of including a larger set of covariates. To further address concerns of over-fitting, I focus on results in which I penalize the inclusion of additional variables. I use rigorous square root lasso (Belloni et al., 2011, 2012) to select the variables to include in the model and then post-estimation OLS (Belloni et al., 2013) to estimate the counterfactual outcomes.[32]

I estimate year-specific effects of triplicate status. In addition, I present more aggregated estimates: 1996-2000, 2001-2010, and 2011-2018. I also present an average treatment effect for 1996-2018. The three sub-period estimates are chosen to summarize effects for

[31] I only allow year variation instead of month since I only report year-specific estimates (or more aggregated metrics) so there is little loss in only accounting for confounders at this level. The benefit of not permitting month-level parameter variation is reducing the number of parameters in the model and concerns about over-fitting.

[32] This method is implemented using `lassopack` in Stata (Ahrens et al., 2019, 2020). I only penalize that interactions of the covariates with year indicators.

the ramp-up of OxyContin's introduction and marketing (1996-2000), the first wave of the opioid crisis (2001-2010), and the post-reformulation transition to illicit opioids (2011-2018). The baseline period is 1981-1995 for all estimates, including the event study estimates. It should be evident from the event study estimates that choice of a different pre-period (e.g., 1991-1995) would not impact the main results of this paper.

For all estimates I produce 95% confidence intervals using the suggested approach. This approach requires testing a series of null hypotheses – the lowest and highest values that are not rejected define the confidence interval. These confidence intervals are not necessarily symmetric.

4.3 Trends in Covariates

Figure 1 above suggested that triplicate and non-triplicate states had different trends prior to 1996. However, these differential trends can be explained by a few demographic shifts. I explore these in detail in Appendix Section C. The goal of this exercise is to find variables which typically predict labor force participation but also appear to be differentially changing by triplicate status. I find evidence that the share white and non-Hispanic, share Hispanic, share with a high education (defined as some college or more), and share ages 45-54 have differential trends across treatment.

To summarize the evidence, I illustrate the implications on labor supply. I estimate the relationship between these covariates and the labor force participation rate prior to 1996. I use the estimated coefficients to construct predicted labor force participation rates for the full sample period. This exercise helps illustrate the main method of this paper. Figure 3 presents the observed differences in labor force participation rates between triplicate and non-triplicate states as well as the predicted differences based only on observable characteristics. There is a steep relative trend in labor force participation, but this trend can be explained by demographic changes.

Notably, we observe large post-treatment differences in labor force participation that cannot be explained by trends in observables. These differences will be attributed to treatment in our main analysis.

5 Results

5.1 Mortality Effects

Before proceeding to the labor supply analysis, I first examine the differential impact of OxyContin's launch on overdose deaths. Alpert et al. (2021a) finds large differences in overdose rate growth post-1996 between triplicate and non-triplicate states, relying primarily on event studies without covariates, making the econometric concerns discussed in this paper irrelevant.[33] Here, I replicate the overdose rate analysis of Alpert et al. (2021a) but first residualize the outcomes using the covariates from this paper.[34] The results are presented in Figure A5 and are similar to the unadjusted results found in earlier work. While the mortality effects are large, they are small relative to the labor supply effects found below (i.e., mortality is a rare event while working is not). I ignore potential selection effects due to mortality given the order of magnitude difference in the effect sizes.

5.2 Modified Event Studies

I present the main results in Figure 4. There is no evidence of any pre-existing trends. Beginning soon after OxyContin's launch, I estimate differential growth in labor supply, which continues almost continuously through the end of the sample period. This pattern holds regardless of whether the covariates are permitted to have different effects in each year (Panels B and D).

I summarize these results in Table 1. In Column 1, I residualize using only the baseline covariates with no year-specific interactions. Column 2 interacts all covariates with year indicators. Due to concerns about over-fitting, I use lasso in Column 3 to select the covariates interacted with year indicators. The results are similar across all columns. Focusing on Column 3, I estimate a small (and insignificant) increase in labor force participation for 1996-2000. The estimates increase over time. For the 2011-2018 period, I estimate that triplicate states experienced relative growth in labor force participation of 3.0 percentage points, statistically significant from zero at the 1% level. For the full post-period, I find that triplicate states had an additional 1.9 percentage points of labor force participation, on average, due to its reduced exposure to OxyContin. For the share working, I observe even larger estimates, including an average effect of 3.8 percentage points.

[33] Results with covariates are included in the appendix and Table 1 of Alpert et al. (2021a). There is some evidence of attenuation when covariates are included additively. However, the attenuation appears to be less important in the context of overdose deaths than when estimating labor supply effects.

[34] Similar analysis with a focus on the 25-54 population generates similar (but larger in magnitude) results.

These estimates are large and the lower end of the confidence intervals exclude economically meaningful rates of growth. For labor force participation, the 95% confidence intervals excludes values lower than 0.7 while I can reject that the percentage of people working grew by less than 2.3 percentage points. I discuss below that these estimates are at the lower end of those found in the literature.

The event study estimates suggest rather continuous differential growth of triplicate state labor supply relative to non-triplicate state labor supply, which is consistent with the continuous differential growth in OxyContin (and oxycodone) access. The timing of the labor supply effect appears to be sooner than the overdose death effect (though there are small mortality differences beginning in 1996), which would be consistent with a rise in misuse preceding a rise in overdose deaths. The labor supply effects continue even after reformulation in 2010, consistent with findings in Park and Powell (2021) and Cho et al. (2021) about the role of illicit markets in reducing labor supply in states more exposed to OxyContin.

5.3 Traditional Difference-in-Differences Estimates

For comparison purposes, I estimate traditional event studies for both labor force participation and share working. I regress the outcome on state fixed effects, time fixed effects, and triplicate-year interactions. I show results without additional covariates in Figure A6. When I condition on additional variables, I adopt the traditional approach of including them additively. I plot the estimates on the triplicate-year interactions in Figure A7. The 1995 interaction is normalized to zero. 95% confidence intervals are generated using the Ferman and Pinto (2019) approach. The equivalent difference-in-differences estimates are presented in Table A1. Interestingly, the overall (1996-2018) estimate is *negative* and statistically significant from zero in the preferred specification.[35]

Comparing Figure A7 with Figure 4 suggests that treatment heterogeneity is substantially biasing the pre-treatment estimates in the traditional event study approach. Adopting a more flexible approach eliminates these pre-existing trends. The difference-in-differences results are especially sensitive to treatment heterogeneity as well, illustrating the importance of residualizing prior to estimating the treatment effects.

Given the pre-trends observed when shifting demographic are not accounted for (see

[35]The difference-in-differences estimates are especially sensitive to the definition of the pre-period, which should be clear from Figure A6. Alternative calculations of the average treatment effect, such as estimating an event study and then averaging the year-specific estimates partially alleviates the bias shown here.

Figure A6), researchers might typically try to address them by estimating a model with state-specific (or treatment-specific) trends or using alternative approaches like synthetic control estimation (Abadie et al., 2010), among other options. Methods which use pre-treatment trends to forecast post-treatment outcomes do not necessarily adequately account for post-treatment outcomes changes due to covariates. For example, consider a covariate which has a differential trend prior to treatment in the treated states relative to the untreated states, but this trend stops after treatment. The counterfactual outcome may also have a different trend, induced by this covariate, but we would not expect the outcome trend to continue throughout the post-period.

5.4 Robustness Tests

I consider possible confounders for the results provided in Section 5.2 and present the results in Table 2. In the first column, I use a larger set of covariates. I include the share non-Hispanic and Black (in addition to the previous race/ethnicity variables), four education share variables, and three age group shares. These are also interacted by treatment status and by time indicators in the post-period. The estimates are generally larger in magnitude when I include this fuller set of controls to estimate the counterfactual outcomes. Because of concerns of over-fitting, I also provide results in which inclusion of additional variables is penalized using the same method as before. The results are similar to the main estimates.

The triplicate states are notably larger in population size relative to most non-triplicate states. In Column 3, I add the log of population size to the set of covariates (also interacted with treatment status-gender and year-gender dummies). Controlling for population size has little effect on the estimates. Column 4 repeats Column 3 but uses lasso, as before, to select variables in the model. Results are similar.[36]

In Column 5, I include a set of policy variables to address possible confounding policy adoption. These policy controls include any PDMP, an electronic PDMP, medical marijuana law, legal and operational medical marijuana dispensaries, the state earned income tax credit (EITC) rate (as a function of the federal rate), and the log of the minimum wage.[37] The inclusion of these variables has little effect on the results.

A related concern is that triplicate states were ahead of the curve in deterring

[36]There may be concerns of nonlinear (log of) population effects, but it is notable the estimates are nearly-identical when permitting large population states to have differential growth in labor outcomes.

[37]The PDMP and medical marijuana policy data are from OPTIC (RAND OPTIC, 2021a,b). EITC and minimum wage information are from University of Kentucky Center for Poverty Research (2021).

opioid misuse or substance use more generally and would have experienced different labor supply trajectories even in the absence of OxyContin. At the time of OxyContin's launch, several states had PDMPs, mostly electronic programs, which we would expect to also be at the frontier of substance use prevention, suggesting that these states may provide more appropriate counterfactuals. I limit the analysis to states with PDMPs in 1996 in Column 6. The estimates are larger than the main estimates.

In Column 7, I exclude the ASEC sample and, in Column 8, I use only the ASEC sample. Results are consistent across the two analysis samples. Finally, in Column 9, I study the log of the labor force participation rate and the log of the share working. The point estimates imply similar level effects as the main estimates, suggesting that the results are not driven by functional form assumptions.

5.4.1 Labor Demand Shocks

Additionally, I test whether the main estimates are sensitive to accounting for (exogenous) labor demand shocks. I present these results in Table A2. The main results are repeated in the first column. Betz and Jones (2018) find that overdose death rates respond to labor demand shocks using a Bartik-style instrument (Bartik, 1991). I construct a similar instrument, predicting the share working by interacting baseline (1995) state-specific industry shares with national (by subtracting out each state's own growth) industry-level growth. The results are similar when I include this control. Charles et al. (2019) consider the role of the decline in manufacturing employment in the United States and also find a relationship with overdose deaths. I construct a similar Bartik-type instrument for manufacturing employment specifically and control for this measure as well. The results are generally unaffected.

Finally, Pierce and Schott (2020) find that trade liberalization policy has had differential geographic effects on "deaths of despair." I control for their measure of industry exposure to permanent normal trade relations to China,[38] interacted with year dummies. These estimates are in the last column of Table A2. In general, as I include additional measures of labor demand, the estimates increase in magnitude.

5.4.2 Advancements in the Returns to Leisure

Aguiar et al. (2021) conclude that improvements in video gaming have reduced the labor supply of men ages 21-30.[39] In this section, I consider whether this trend could be

[38]Data found at `https://www.aeaweb.org/articles?id=10.1257/aeri.20180396` and aggregated to the state level.

[39]See Kimbrough (2019) for further discussion.

differentially impacting triplicate and non-triplicate states. Following Aguiar et al. (2021), I use American Time Use Data (ATUS) (Hofferth et al., 2020) to study the number of minutes per day respondents were engaged in "Playing Games." Since my interest is in understanding the effects on the 25-54 population, I select on this population. I do not observe a pre-period for this outcome so all results reflect adjusted cross-sectional differences. The residualization step only involves year-gender fixed effects and covariates interacted with year-gender dummies. These parameters are estimated using the non-triplicate states only.

Figure A8, Panel A provides the unadjusted trends. Panels B and C show adjusted differences. Overall, there is little evidence of meaningful differences between triplicate and non-triplicate states. The magnitudes of the unadjusted differences are never larger than 2 minutes per day (and smaller in the adjusted differences). Below, I find that the difference in hours worked per week far surpasses this magnitude.

5.4.3 Migration

It is important to consider migration for two reasons. First, the results may be driven by selective migration. People with different working propensities may select to live in places with more opioid access or avoid them because of increased crime due to the development of illicit drug markets. The implications of the findings above are different if they are due to systematic differences in migration. Second, the empirical strategy of this paper is focused on the role of covariates, and I show that these covariates are important in explaining differences across treated and untreated states. The assumption is that the covariates are exogenous and not responding to OxyContin exposure. Migration could undermine this assumption.[40]

I test this assumption by studying migration and immigration directly since the CPS reports residence changes in the past year. I study total in-migration (scaled by state population) (Panel A of Figure A9), total out-migration (scaled by state population) (Panel B of Figure A9), composition of individuals migrating into the state (Figure A10), and composition of individuals migrating out of the state (Figure A11). For composition, I study the four covariates discussed in Section 4.3. I find no evidence of systematic migration on any dimension.

[40]We would not expect any endogenous changes in birth rates to have much of an impact on age 25-54 labor outcomes during this time period.

5.5 Other Measures of Labor Supply

In this section, I study alternative measures of labor supply. First, I use data from the Current Employment Statistics (CES) to construct an alternative measure of the employment-to-population ratio which does not rely on self-reported information. The CES is a survey of establishments representing workers covered by unemployment insurance. Each month, the CES surveys about 145,000 nonfarm businesses and government agencies. By the nature of the survey design, the CES excludes some industries and the self-employed. It is designed to provide the number of jobs based on place of work, not place of residence. I scale these employment figures by the total resident population ages 16 and above with the understanding that people may reside in one state and work in another. While there are limitations to the CES, it offers a useful complementary, establishment-level metric of employment. The CES estimates are provided in Figure 5, Panel A. The pattern of the estimates is similar to the CPS results, though the magnitudes tend to be larger.

In Panel B, I provide the corresponding results using BEA data. The BEA builds on the CES while including information about the self-employed.[41] The results are similar to the CES results.

Next, I study labor supply measures which also incorporate intensive margin decisions (in addition to the decision to work or not). In Panel C, I study "hours worked last week," equal to zero for non-workers, in the CPS. This variable is available beginning in 1989 for all samples. In Panel D, I study the log of annual labor earnings, reported in the ASEC. For both hours worked and annual labor earnings, I observe evidence of large effects, consistent with the main results of the paper.

As a complementary measure, I use state-level data from the Bureau of Economic Analysis (BEA) on total employee compensation divided by total population ages 16+. The BEA compensation metric includes wages and salaries as well as the value of noncash benefits (e.g., employer contributions to health insurance and pension plans). I present these results in Panel E. The pattern of results are generally similar to the patterns observed for the other labor outcomes in this paper.

5.6 Heterogeneity

With only five treated units, it is difficult to isolate the sources of heterogeneity biasing the traditional difference-in-differences results. In this section, I explore heterogeneity

[41] The BEA collects information from Internal Revenue Service data to estimate sole proprietorships and nonfarm partners such that the final employment numbers include the self-employed.

based on demographics, presented in Table A3. I first present estimates by gender. The point estimates in Table A3 tend to be larger for women than men, suggesting that female labor supply was more responsive to OxyContin exposure, though men also experienced large labor supply effects. The relative magnitudes parallel those found by Krueger (2017).

The opioid crisis, as measured by overdose deaths, has disproportionately affected men, though the effects have been substantial for women too (Singh et al., 2019), especially relative to previous drug epidemics. These labor supply effects potentially reflect that while women are less likely to die from opioid overdoses, they might be more affected on other margins. For example, women may be forced to take on more caretaking duties given a rise in opioid dependence within the household, reducing working opportunities. Women may also be more responsive to the broader economic effects of the opioid crisis, consistent with much of the female labor supply literature suggesting a high level of responsiveness to after-tax wages (Keane, 2011).

Next, I study effects by race/ethnicity. I observe the largest effects for the white and non-Hispanic population. Finally, I stratify the sample based on education status. The effect sizes are largest for the low education group. These last results are more consistent with the mortality effects of the opioid crisis.

The main analysis centered on the 25-54 population given the relatively high rates of labor force attachment for this group. I study age-based heterogeneity in this section. First, I estimate effects for the 18-64 population. Table A4 presents the results. The estimated effects are comparable, though smaller, to those observed for the 25-54 population. I estimate statistically significant effects for the full post-period of 2.0 and 2.7 percentage points for labor force participation and working, respectively.

In the next 3 columns, I study more disaggregated age groups: ages 25-44, ages 45-64, and ages 65+. One motivation for this analysis is the possibility that additional access to opioid therapy in non-triplicate states may have improved outcomes for older age groups given the higher incidence of pain-related work-limiting disability among this population. The effect sizes for the 65+ age group are negative for 1996-2000 and 2001-2010, though very small in magnitude (and not statistically different from zero). The small effect sizes are potentially due to the reduced scope for affecting labor supply. They may also reflect that this population does benefit from the increased opioid access, in terms of employment propensities, while also suffering from additional levels of misuse. These two effects may be cancelling each other out.

The biggest effects are observed for the 25-44 age group. I also estimate positive and large effects for the 45-64 age group. The age patterns are consistent with the age groups most impacted by the opioid crisis in terms of overdose deaths (Hedegaard et al., 2020).

Finally, I study employment by industry. I present the average (over 1996-2018) effects on employment in industries divided by population size (for ages 25-54). The estimates are provided in Figure A12. It is hard to predict which industries would be disproportionately impacted by opioid access. Jobs with higher injury rates may benefit from the additional pain management therapy access. However, workers in these jobs may be more likely to misuse opioids due to the additional access.[42] I observe positive and statistically significant employment effects in five industries. I estimate negative effects for some industries, including a large negative effect for manufacturing, but none are statistically different from zero.

5.7 Discussion of Magnitudes

The estimated effects on labor force participation and working rates are large. For comparison, Krueger (2017) estimates that each 10% increase in opioid prescriptions decreases labor force participation by 0.11 percentage points for men and 0.14 for women. Using a similar empirical strategy, Aliprantis et al. (2019) conclude that each 10% increase in opioid prescriptions decreases labor force participation of working age men by between 0.15 and 0.47 percentage points. For working age women, the decrease is between 0.15 and 0.19. Beheshti (2019) concludes that a 10% reduction in hydrocodone prescriptions leads to a 0.20 percentage point increase in labor force participation.

The contexts and identification sources of these estimates are different than ours. Krueger (2017) and Aliprantis et al. (2019) assume that changes in local opioid supply are exogenous to economics conditions. The rescheduling of hydrocodone, analyzed in Beheshti (2019), occurred during a time when illicit opioid markets had developed (Powell and Pacula, 2021). We likely expect that the large growth of opioid access in the first wave would have different effects than declines in prescribing during the fentanyl crisis.

As one measure of differential exposure used in Alpert et al. (2021a), I use the differences in initial OxyContin supply, measured in 2000, the first year available from ARCOS but also conveniently reflecting a useful baseline given that it took a few years to reach an initial steady-state. In 2000, non-triplicate states had 1.14 MEDs per capita of OxyContin compared to 0.43 in triplicate states, implying 165% higher exposure in non-triplicate

[42] Musse (2020) provides industry-specific analyses which more directly model these mechanisms.

states.[43]

Using the Table 1, Column 3 estimate for 1996-2018, the implied relationship is that each 10% increase in exposure to OxyContin decreased labor force participation by 0.12 percentage points. As an alternative metric I use growth in oxycodone supply.[44] The ARCOS data only begin in 1997 so I compare average per capita oxycodone MEDs to this 1997 "baseline," recognizing that 1997 was also treated and includes some of the effect on oxycodone supply. Per capita oxycodone MEDs increased, on average, by 3.34 in non-triplicate states and 1.48 in triplicate states. Thus, the growth was 125% higher in non-triplicate states. The implied relationship between oxycodone supply and labor force participation is that each 10% increase in oxycodone supply decreases labor force participation by 0.15 percentage points.

These estimates are at the low end of those found in the literature. While the overall results appear large, this is a function of the substantial geographic variation in the shock to OxyContin and oxycodone supply, not because I am estimating abnormally large labor force participation responses to changes in opioid supply.

6 Conclusion

This paper finds evidence that the opioid crisis led to large reductions in labor supply in the United States. These reductions began soon after the introduction of OxyContin and grew in magnitude throughout the sample period, including after the removal of the original formulation of OxyContin, consistent with prior evidence found in Park and Powell (2021). Notably, traditional difference-in-differences estimates, conditioning on a small set of covariates, obscure this result. This paper shows how including covariates additively in an event study or difference-in-differences specification can lead to bias in the presence of heterogeneous treatment effects.

There is interest in understanding the role of the opioid crisis in the national reduction of labor supply. Krueger (2017) focuses on the reduction from 1999 to 2015. During this time period, labor force participation rates dropped by 3.2 percentage points. Notably, demographic changes – a hypothesized explanatory factor in the literature – predict no change

[43] "Exposure" here refers broadly to early access, promotional activity, and subsequent spillovers to the prescribing of other opioids resulting from initial differences.

[44] Using oxycodone has the advantage that I have data back to 1997 (instead of 2000 for OxyContin). The oxycodone measure also directly includes spillovers to other oxycodone products due to Purdue Pharma's promotional activities (though the OxyContin exposure measure implicitly incorporates the resulting spillovers increase oxycodone use more broadly).

in the labor force participation rate over this time[45] in my model as the rise in education rates counteracted declines due to population aging (within the 25-54 age group). My event study estimates imply that triplicate states experienced a different increase in labor force participation of 2.3 percentage points during this time period.

I perform the same extrapolation exercise provided in Alpert et al. (2021a) with the usual caveats about out-of-sample predictions. One of these standard caveats is that the estimated effect on the treated units is not necessarily the same effect that would have been experienced by the untreated states. I highlight this concern given this paper's focus on heterogeneity.[46] Alpert et al. (2021a) noted that the differential effects estimated across states refer to differences in initial OxyContin exposure, which I proxy using the first year of available ARCOS data. This difference is equal to 0.71 morphine equivalent doses (MEDs). I extrapolate to estimate the impact of the national-level of initial exposure to OxyContin, equal to 0.92 MEDs.

This extrapolation exercise concludes that differential initial exposure to OxyContin accounts for a 2.6 percentage point reduction in labor force participation from 2000 to 2015, a substantial share of the total change in participation rates at this time. However, it is worth noting that labor force participation was generally trending upwards prior to the opioid crisis, suggesting that a full census of possible reasons for the decline in labor force participation could add up to more than 100% (given a counterfactual in which labor force participation rates continued to rise).

The large labor supply consequences of the opioid crisis reveal the broader impacts of this epidemic, suggesting that we should expect that other dimensions of individual and household life may also have been affected. The opioid crisis is potentially a first-order factor when evaluating a myriad of metrics related to individual and household welfare. Using variation in initial OxyContin exposure is a useful approach for uncovering these consequences. However, this paper also shows that it is critical to account for changing demographic factors across triplicate and non-triplicate states and do so in a manner which does not induce bias into the estimates. One such approach is suggested in this paper. The analysis of overdose deaths is robust to these concerns; however, preliminary analysis on

[45] I estimate that demographic changes *increased* labor force participation rates by 0.03 percentage points from 1999 to 2015.

[46] Of course, it may be that there is substantial treatment heterogeneity among triplicate states but that, on average, treatment would have had the same impact in non-triplicate states. This type of assumption is typically necessary to predict national impacts from sub-national analyses.

other outcomes – such as household composition – suggests that the bias documented for labor supply extends to other contexts as well.

References

Aaronson, Daniel, Jonathan Davis, and Luojia Hu, "Explaining the decline in the US labor force participation rate," *Chicago Fed Letter*, 2012, (296), 1.

Abadie, Alberto, "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

_ , Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

Abraham, Katharine G and Melissa S Kearney, "Explaining the decline in the US employment-to-population ratio: A review of the evidence," Technical Report, National Bureau of Economic Research 2018.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst, "Leisure luxuries and the labor supply of young men," *Journal of Political Economy*, 2021, *129* (2), 000–000.

Ahrens, Achim, Christian B Hansen, and Mark E Schaffer, "lassopack: Model selection and prediction with regularized regression in Stata," *The Stata Journal*, 2020, *20* (1), 176–235.

_ , _ , and Mark Schaffer, "LASSOPACK: Stata module for lasso, square-root lasso, elastic net, ridge, adaptive lasso estimation and cross-validation," 2019.

Alexander, Monica J, Mathew V Kiang, and Magali Barbieri, "Trends in black and white opioid mortality in the United States, 1979–2015," *Epidemiology (Cambridge, Mass.)*, 2018, *29* (5), 707.

Aliprantis, Dionissi, Kyle Fee, and Mark E Schweitzer, "Opioids and the labor market," 2019.

Alpert, Abby, David Powell, and Rosalie Liccardo Pacula, "Supply-side drug policy in the presence of substitutes: Evidence from the introduction of abuse-deterrent opioids," *American Economic Journal: Economic Policy*, 2018, *10* (4), 1–35.

Alpert, Abby E, William N Evans, Ethan MJ Lieber, and David Powell, "Origins of the Opioid Crisis and Its Enduring Impacts," Technical Report, National Bureau of Economic Research 2021.

Alpert, Abby, Stephen Schwab, and Ben Ukert, "Opioid Abuse and Employment Outcomes: Evidence from the US Military," Technical Report 2021.

Andersen, Martin, Johanna Catherine Maclean, Michael F Pesko, and Kosali I Simon, "Effect of a Federal Paid Sick Leave Mandate on Working and Staying at Home: Evidence from Cellular Device Data," Technical Report, National Bureau of Economic Research 2020.

Athey, Susan and Guido W Imbens, "Design-based analysis in difference-in-differences settings with staggered adoption," *Journal of Econometrics*, 2021.

Bartik, Timothy J, "Who benefits from state and local economic development policies?," 1991.

Beheshti, David, "The Impact of Opioids on the Labor Market: Evidence from Drug Rescheduling," 2019.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen, "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 2012, *80* (6), 2369–2429.

_ , Victor Chernozhukov, and Lie Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, 2011, *98* (4), 791–806.

_ , _ et al., "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 2013, *19* (2), 521–547.

Betz, Michael R and Lauren E Jones, "Wage and employment growth in America's drug epidemic: Is all growth created equal?," *American Journal of Agricultural Economics*, 2018, *100* (5), 1357–1374.

Bilinski, Alyssa and Laura A Hatfield, "Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions," 2020.

Black, Sandra, Jason Furman, Emma Rackstraw, and Nirupama Rao, "The long-term decline in US prime-age male labour force participation," *VoxEU.org*, 2016.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and John Van Reenen, "Evaluating the employment impact of a mandatory job search program," *Journal of the European Economic Association*, 2004, *2* (4), 569–606.

Borusyak, Kirill and Xavier Jaravel, "Revisiting event study designs," *Available at SSRN 2826228*, 2017.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Available at SSRN 3148250*, 2019.

Cameron, A Colin and Pravin K Trivedi, *Microeconometrics: methods and applications*, Cambridge university press, 2005.

Canay, Ivan A, Andres Santos, and Azeem M Shaikh, "The wild bootstrap with a "small" number of "large" clusters," *Review of Economics and Statistics*, 2019, pp. 1–45.

_ , Joseph P Romano, and Azeem M Shaikh, "Randomization tests under an approximate symmetry assumption," *Econometrica*, 2017, *85* (3), 1013–1030.

Carpenter, Christopher S, Chandler B McClellan, and Daniel I Rees, "Economic conditions, illicit drug use, and substance use disorders in the United States," *Journal of Health Economics*, 2017, *52*, 63–73.

Case, Anne and Angus Deaton, "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences*, 2015, *112* (49), 15078–15083.

_ and _ , "Mortality and morbidity in the 21st century," *Brookings papers on economic activity*, 2017, *2017* (1), 397–476.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz, "The transformation of manufacturing and the decline in US employment," *NBER Macroeconomics Annual*, 2019, *33* (1), 307–372.

Cho, David, Daniel I Garcia, Joshua Montes, and Alison Weingarden, "Labor Market Effects of the Oxycodone-Heroin Epidemic," 2021.

Cicero, Theodore J, James A Inciardi, and Alvaro Muñoz, "Trends in abuse of OxyContin® and other opioid analgesics in the United States: 2002-2004," *The Journal of Pain*, 2005, *6* (10), 662–672.

Commonwealth of Massachusetts, "Commonwealth of Massachusetts v. Purdue Pharma L.P., Purdue Pharma Inc., Richard Sackler, Theresa Sackler, Kathe Sackler, Jonathan Sackler, Mortimer D.A. Sackler, Beverly Sackler, David Sackler, Ilene Sackler Lefcourt, Peter Boer, Paulo Costa, Cecil Pickett, Ralph Snyderman, Judy Lewent, Craig Landau, John Stewart, and Mark Timney," Technical Report 2018.

Council of Economic Advisers, "The Labor Force Participation Rate Since 2007: Causes and Policy Implications.," 2014.

Currie, Janet and Hannes Schwandt, "The Opioid Epidemic Was Not Caused by Economic Distress But by Factors that Could be More Rapidly Addressed," Technical Report, National

Bureau of Economic Research 2020.

_ , **Jonas Jin, and Molly Kathleen Schnell**, "US employment and opioids: Is there a connection?," *Research in Labor Economics*, 2019, *47*, 253–280.

de Chaisemartin, Clément and Xavier D'Haultfoeuille, "Fuzzy differences-in-differences," *The Review of Economic Studies*, 2018, *85* (2), 999–1028.

_ **and** _ , "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Ferman, Bruno and Cristine Pinto, "Inference in differences-in-differences with few treated groups and heteroskedasticity," *Review of Economics and Statistics*, 2019, *101* (3), 452–467.

Fernald, John G, Robert E Hall, James H Stock, and Mark W Watson, "The disappointing recovery of output after 2009," Technical Report, National Bureau of Economic Research 2017.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, and J. Robert Warren, *Integrated Public Use Microdata Series, Current Population Survey: Version 7.0 [dataset]*, Minneapolis: University of Minnesota, 2020.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," Technical Report, National Bureau of Economic Research 2018.

Green, Carla A, Nancy A Perrin, Shannon L Janoff, Cynthia I Campbell, Howard D Chilcoat, and Paul M Coplan, "Assessing the accuracy of opioid overdose and poisoning codes in diagnostic information from electronic health records, claims data, and death records," *Pharmacoepidemiology and drug safety*, 2017, *26* (5), 509–517.

Groups Plus, "Purdue Pharma Company Focus Group Research & Findings Oxycontin for Non-Cancer Pain Management," Technical Report 1995.

Harris, Matthew C, Lawrence M Kessler, Matthew N Murray, and Beth Glenn, "Prescription Opioids and Labor Market Pains: The Effect of Schedule II Opioids on Labor Force Participation and Unemployment," *Journal of Human Resources*, 2019, pp. 1017–9093R2.

Hedegaard, Holly, Arialdi M Miniño, and Margaret Warner, "Drug Overdose Deaths in the United States, 1999–2019," *NCHS Data Brief*, 2020, *394*.

Hofferth, Sandra L., Sarah M. Flood, Matthew Sobek, and Daniel Backman, "American Time Use Survey Data Extract Builder: Version 2.8 [dataset]," 2020.

Hollingsworth, Alex, Christopher J Ruhm, and Kosali Simon, "Macroeconomic conditions and opioid abuse," *Journal of Health Economics*, 2017, *56*, 222–233.

Hotchkiss, Julie L and Fernando Rios-Avila, "Identifying factors behind the decline in the US labor force participation rate," *Business and Economic Research*, 2013, *3* (1), 257.

Ibragimov, Rustam and Ulrich K Müller, "Inference with few heterogeneous clusters," *Review of Economics and Statistics*, 2016, *98* (1), 83–96.

Jaeger, David A, Theodore J Joyce, and Robert Kaestner, "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?," *Journal of Business & Economic Statistics*, 2018, pp. 1–10.

Juhn, Chinhui and Simon Potter, "Changes in labor force participation in the United States," *Journal of Economic Perspectives*, 2006, *20* (3), 27–46.

Kaestner, Robert and Engy Ziedan, "Mortality and socioeconomic consequences of prescription opioids: Evidence from state policies," Technical Report, National Bureau of Economic Research 2019.

Kahn-Lang, Ariella and Kevin Lang, "The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications," *Journal of Business & Economic Statistics*, 2019, pp. 1–14.

Keane, Michael P, "Labor supply and taxes: A survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

Kearney, Melissa S and Phillip B Levine, "Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing," *American Economic Review*, 2015, *105* (12), 3597–3632.

Kimbrough, Gray, "Xboxes and Ex-workers? Gaming and Labor Supply of Young Adults in the US," 2019.

Krueger, Alan B, "Where have all the workers gone? An inquiry into the decline of the US labor force participation rate," *Brookings Papers on Economic Activity*, 2017, *2017* (2), 1.

Laird, Jessica and Torben Nielsen, "The Effects of Physician Prescribing Behaviors on Prescription Drug Use and Labor Supply: Evidence from Movers in Denmark," *Job market paper. http://scholar. harvard. edu/files/lairdja/files/Laird_JMP. pdf*, 2016.

Maclean, Catherine, Justine Mallatt, Christopher J Ruhm, and Kosali Ilayperuma Simon, "Review of Economic Studies on the Opioid Crisis," *NBER Working Paper*, 2020, (w28067).

Musse, Isabel, "Employment Shocks and Demand for Pain Medication," *Available at SSRN 3646543*, 2020.

Negi, Akanksha and Jeffrey M Wooldridge, "Revisiting regression adjustment in experiments with heterogeneous treatment effects," *Econometric Reviews*, 2020, pp. 1–31.

Park, Sujeong and David Powell, "Is the rise in illicit opioids affecting labor supply and disability claiming rates?," *Journal of Health Economics*, 2021, *76*, 102430.

Pierce, Justin R and Peter K Schott, "Trade liberalization and mortality: evidence from US counties," *American Economic Review: Insights*, 2020, *2* (1), 47–64.

Polivka, Anne E and Stephen M Miller, "The CPS after the redesign: Refocusing the economic lens," in "Labor Statistics Measurement Issues," University of Chicago Press, 1998, pp. 249–289.

Powell, David and Rosalie Liccardo Pacula, "The evolving consequences of OxyContin reformulation on drug overdoses," *American Journal of Health Economics*, 2021, *7* (1), 41–67.

Purdue Pharma, "OxyContin Launch Plan," Technical Report 1995.

_ , "1999 Budget Plan," Technical Report 1999.

RAND OPTIC, "OPTIC-Vetted Medical Marijuana Policy Data," Technical Report 2021.

_ , "OPTIC-Vetted PDMP Policy Data," Technical Report 2021.

Roth, Jonathan and Ashesh Rambachan, "An Honest Approach to Parallel Trends," 2019.

Ruhm, Christopher J, "Drivers of the fatal drug epidemic," *Journal of Health Economics*, 2019, *64*, 25–42.

_ , "Nonopioid overdose death rates rose almost as fast as those involving opioids, 1999–2016," *Health Affairs*, 2019, *38* (7), 1216–1224.

Sant'Anna, Pedro HC and Jun Zhao, "Doubly robust difference-in-differences estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

Savych, Bogdan, David Neumark, and Randall Lea, "Do opioids help injured workers recover and get back to work? The impact of opioid prescriptions on duration of temporary disability," *Industrial Relations: A Journal of Economy and Society*, 2019, *58* (4), 549–590.

Singh, Gopal K, Isaac E Kim, Mehrete Girmay, Chrisp Perry, Gem P Daus, Ivy P Vedamuthu, Andrew A De Los Reyes, Christine T Ramey, Elijah K Martin, and Michelle Allender, "Opioid epidemic in the United States: empirical trends, and a literature review of social determinants and epidemiological, pain management, and treatment patterns," *International Journal of Maternal and Child Health and AIDS*, 2019, *8* (2), 89.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2020.

Thingholm, Peter, "Labor Market Effects on Opioid Use: Evidence from Provider Dynamics," Technical Report 2020.

University of Kentucky Center for Poverty Research, "UKCPR National Welfare Data, 1980-2019," Technical Report 2021.

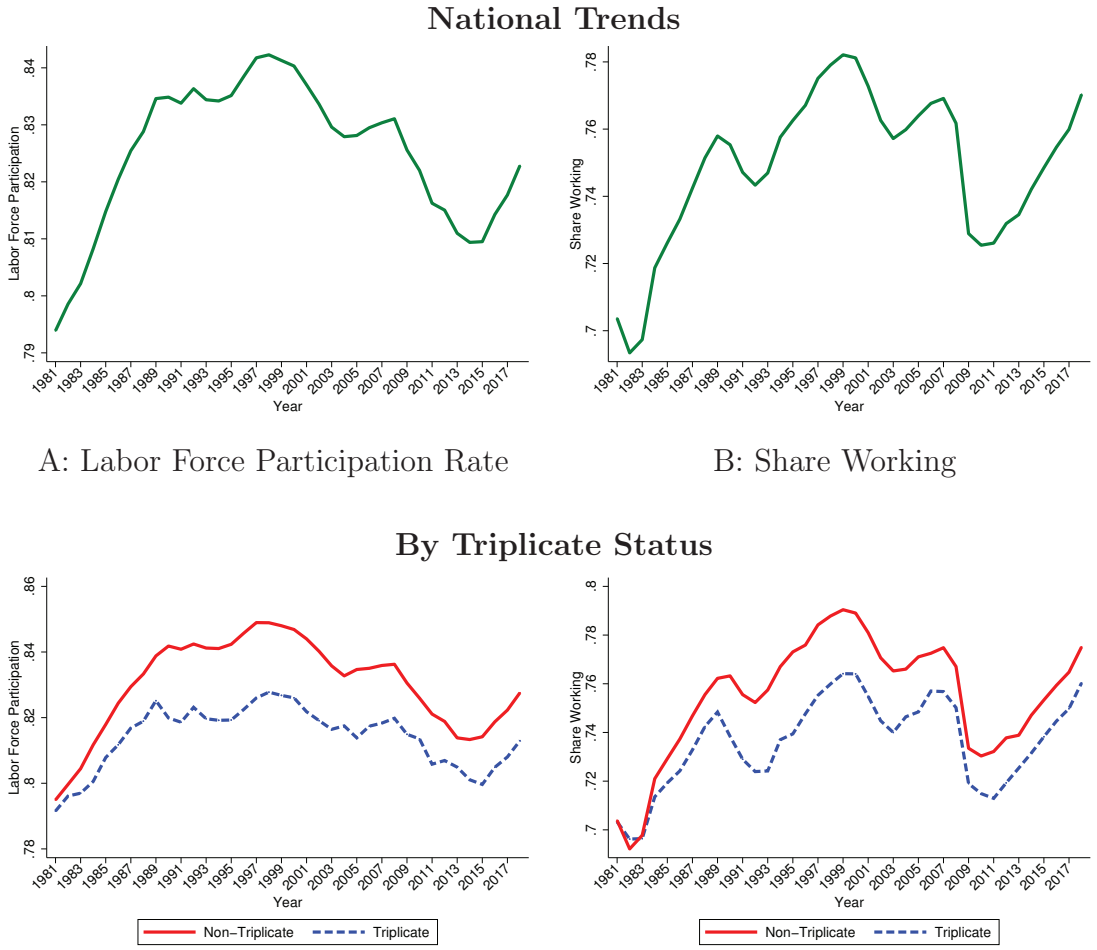
Venkataramani, Atheendar S, Elizabeth F Bair, Rourke L O'Brien, and Alexander C Tsai, "Association between automotive assembly plant closures and opioid overdose mortality in the United States: a difference-in-differences analysis," *JAMA Internal Medicine*, 2020, *180* (2), 254–262.

Warner, Margaret, Li Hui Chen, Diane M Makuc, Robert N Anderson, and Arialdi M Miniño, "Drug poisoning deaths in the United States, 1980–2008," *NCHS Data Brief*, 2011, *81* (81), 1–8.

Wilson, Nana, "Drug and Opioid-Involved Overdose Deaths – United States, 2017–2018," *MMWR. Morbidity and Mortality Weekly Report*, 2020, *69.*

Figures

Figure 1: Labor Supply Trends for 1981-2018, Ages 25-54



A: Labor Force Participation Rate

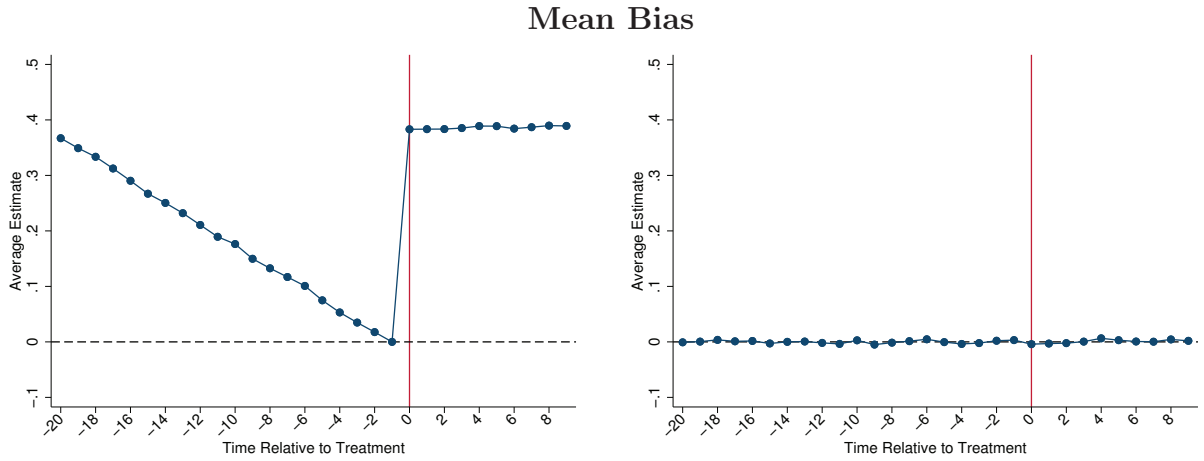
B: Share Working

C: Labor Force Participation Rate

D: Share Working

Notes: I use CPS data, selecting on ages 25-54, to construct the annual labor force participation rate and fraction of the population employed.

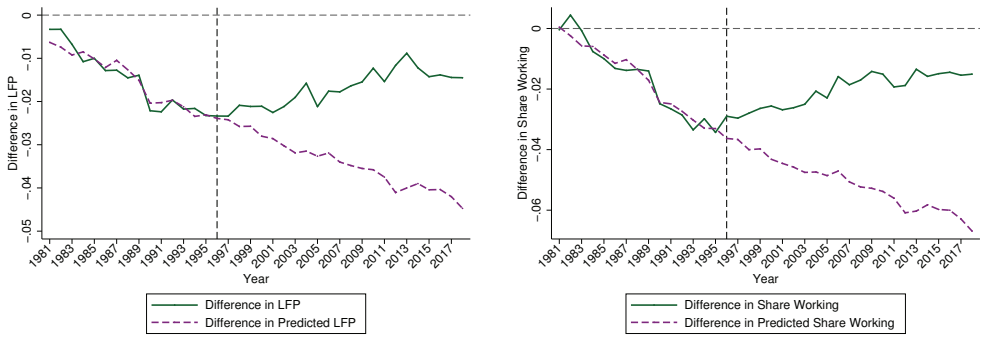
Figure 2: Simulation Results



A: Traditional Event Study B: Modified Event Study

Notes: I plot the mean bias estimates for each time period from 10,000 simulations. See Sections 3.1 and 3.3 for details.

Figure 3: Outcome Differences Between Triplicate and Non-Triplicate States

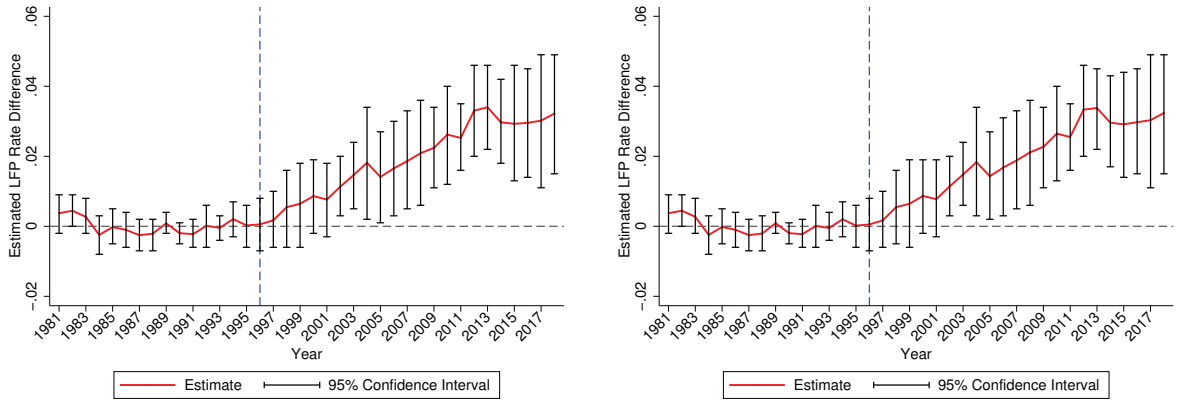


A: Labor Force Participation B: Working

Notes: Each figure plots the observed difference in annual labor force participation rate or share working. It also includes a prediction based only on covariates. Using state*sex*month data, I regress the outcome on state-sex dummies, time-sex dummies, and time-varying covariates using pre-period data only. The estimated parameters are then used to predict the outcome in each year. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. The parameters associated with the covariates are permitted to vary by Triplicate status and sex. Regressions are weighted by CPS sample weights.

Figure 4: Main Event Study Results

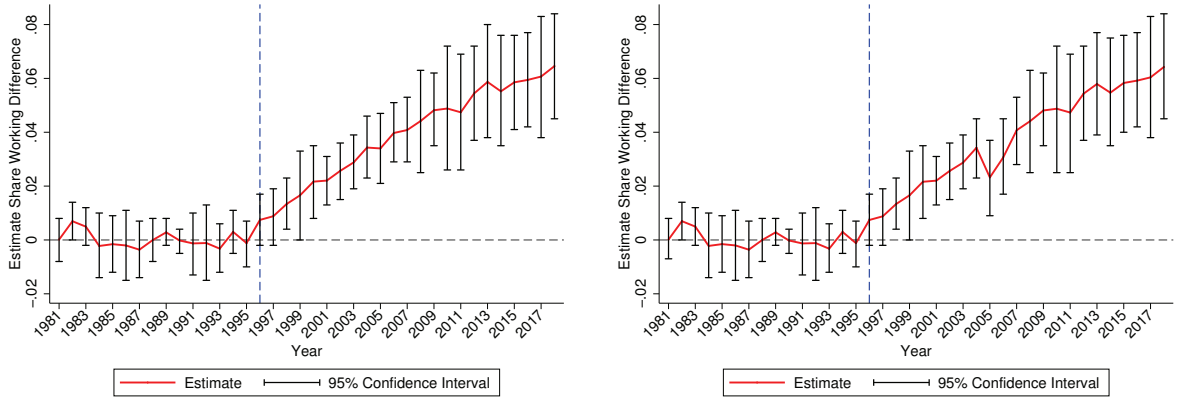
Labor Force Participation Rate



A: Baseline Controls

B: Lasso

Share Working



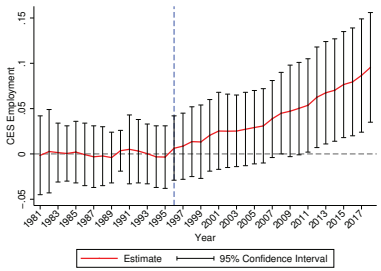
C: Baseline Controls

D: Lasso

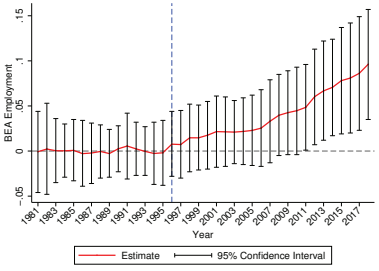
Notes: The outcome is the labor force participation rate (Panels A-B) or the share working (Panels C-D) by month, state, and sex. I regress the outcome on state-sex dummies, time-sex dummies, and time-varying covariates using pre-period data for all states and post-treatment data for non-triplicate states. I residualize the outcome and then present annual (weighted by population) averages of these residuals for the triplicate states. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. Regressions are weighted by CPS sample weights. The parameters associated with the covariates are permitted to vary by sex and Triplicate status. In Panels B and D, the parameters can also vary by gender-year – lasso is used to select these covariates. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure 5: Event Study Estimates for Other Measures of Labor Supply

Number of Jobs Per Ages 16+ Capita

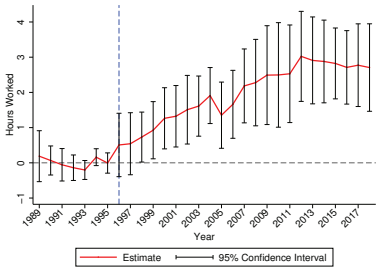


A: Current Employment Statistics

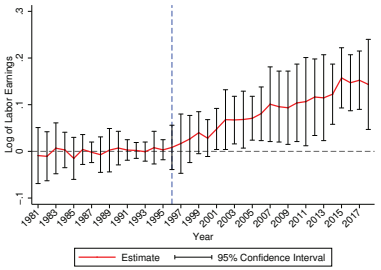


B: Bureau of Economic Analysis

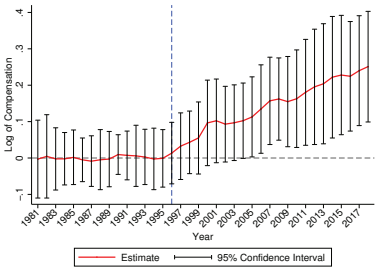
Hours, Earnings, Compensation



C: Weekly Hours Worked (CPS)



D: Log(Annual Earnings) (CPS)



E: Log(Compensation) (BEA)

Notes: In Panels A and B, the outcome is the number of jobs scaled by the population ages 16+. In Panel E, the outcome is the log of per capita (16+) compensation. For these outcomes, I regress the outcome on state dummies, year dummies, and time-varying covariates using pre-period data and post-treatment data for non-triplicate states. I residualize the outcome and then present annual (weighted by 16+ population) averages of these residuals for the triplicate states. Covariates include share (of age 25-54 population) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. Regressions are weighted by population size (16+). The parameters associated with the covariates are permitted to vary by Triplicate status. I also interact the covariates with year dummies and use Lasso to select these covariates. In Panels C and D, the outcomes are weekly hours worked and the log of labor earnings, respectively, in the CPS. Hours worked and labor earnings are set to zero for non-workers before averaging to the state-sex-month level. The estimation procedure is the same except performed at the state-sex-month level so the covariates are permitted to have different effects by sex as well, and the weights are the aggregates of the CPS survey weights. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Tables

Table 1: Modified Difference-in-Differences Estimates

	A: Labor Force Participation		
Triplicate ×	(1)	(2)	(3)
1996-2000	0.005	0.011***	0.005
	[-0.001, 0.011]	[0.004, 0.019]	[-0.002, 0.011]
2001-2010	0.017**	0.021***	0.017**
	[0.004, 0.030]	[0.009, 0.032]	[0.004, 0.031]
2011-2018	0.030***	0.034***	0.030***
	[0.015, 0.046]	[0.022, 0.047]	[0.016, 0.045]
1996-2018	0.019***	0.024***	0.019***
	[0.006, 0.032]	[0.012, 0.035]	[0.007, 0.032]
Covariates	Vary by Treatment	Vary by Time	Selected by Lasso
	B: Working		
Triplicate ×	(4)	(5)	(6)
1996-2000	0.014***	0.018***	0.014***
	[0.006, 0.022]	[0.006, 0.030]	[0.006, 0.022]
2001-2010	0.037***	0.028***	0.035***
	[0.024, 0.049]	[0.013, 0.042]	[0.022, 0.048]
2011-2018	0.057***	0.051***	0.057***
	[0.039, 0.075]	[0.034, 0.069]	[0.039, 0.075]
1996-2018	0.039***	0.034***	0.038***
	[0.024, 0.055]	[0.019, 0.049]	[0.023, 0.054]
Covariates	Vary by Treatment	Vary by Time	Selected by Lasso

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated using the method proposed in Section 3.5. The first step is to residualize outcomes using a regression of the outcome on state-sex and time-sex fixed effects and time-varying covariates. This step uses pre-period from all states and post-treatment data for non-triplicate states. I then report the mean of the residuals over the listed time period for the triplicate states. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. These covariates are permitted to have effects varying by sex. "Vary by Treatment" implies that the covariates are interacted with the Triplicate indicator. "Vary by Time" means that the covariates are also interacted with year dummies. In the final column, I use lasso to select the covariates with parameters varying by time. Regressions and averages are weighted by CPS sample weights.

Table 2: Robustness Tests

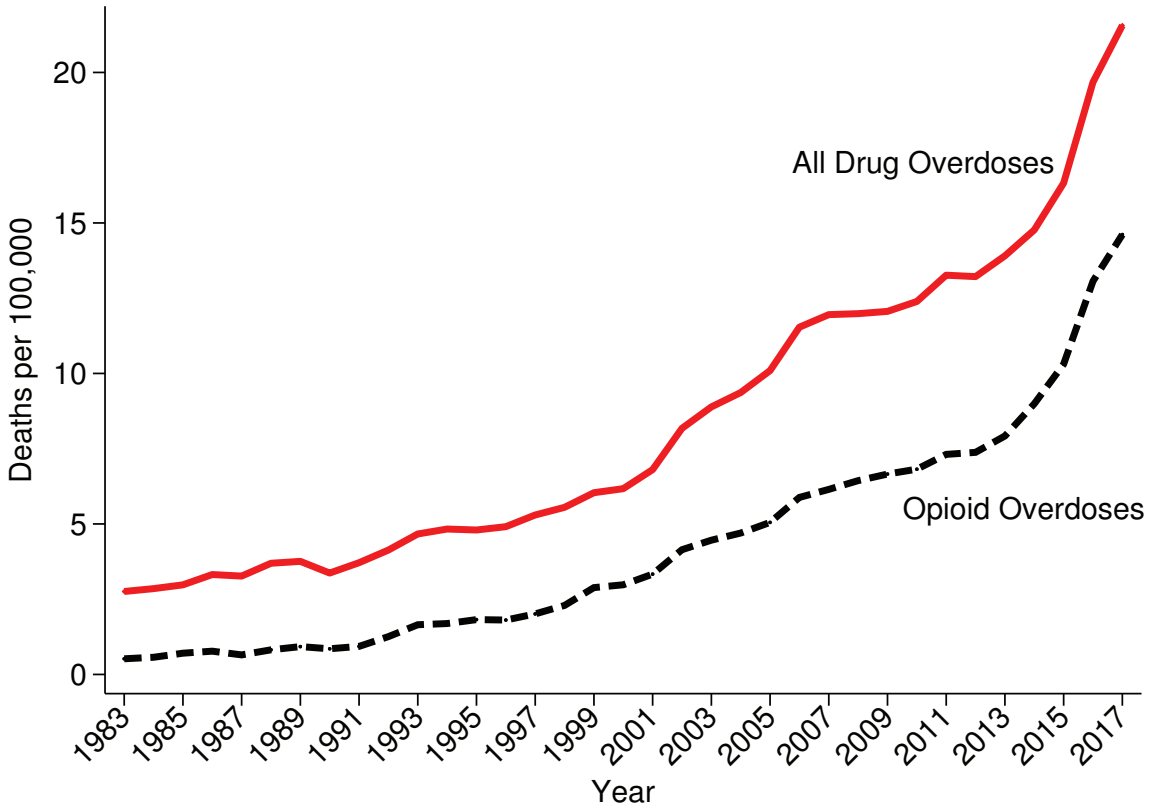
	(1) More Demographic Controls	(2) More Demographic Controls	(3) Add Population	(4) Add Population	(5) Policy Variables	(6) PDMP States	(7) No ASEC	(8) ASEC Only	(9) Log of Outcome
A. Labor Force Participation									
1996-2000	0.018***	0.009**	0.016***	0.008**	0.011***	0.029***	0.011***	0.015**	0.014**
	[0.013, 0.024]	[0.003, 0.015]	[0.010, 0.022]	[0.002, 0.014]	[0.004, 0.018]	[0.015, 0.043]	[0.004, 0.018]	[0.003, 0.027]	[0.001, 0.026]
2001-2010	0.028***	0.025***	0.025***	0.023***	0.018***	0.030***	0.020***	0.021**	0.025***
	[0.020, 0.037]	[0.012, 0.037]	[0.016, 0.034]	[0.014, 0.032]	[0.006, 0.030]	[0.013, 0.046]	[0.009, 0.032]	[0.005, 0.037]	[0.010, 0.040]
2011-2018	0.038***	0.041***	0.034***	0.041***	0.042***	0.046***	0.034***	0.036***	0.042***
	[0.025, 0.050]	[0.026, 0.056]	[0.023, 0.045]	[0.027, 0.055]	[0.028, 0.057]	[0.028, 0.064]	[0.021, 0.047]	[0.014, 0.057]	[0.025, 0.060]
1996-2018	0.030***	0.027***	0.026***	0.026***	0.025***	0.035***	0.023***	0.025**	0.029***
	[0.020, 0.040]	[0.014, 0.040]	[0.018, 0.035]	[0.015, 0.038]	[0.013, 0.037]	[0.020, 0.051]	[0.012, 0.035]	[0.008, 0.043]	[0.013, 0.045]
B. Working									
1996-2000	0.023***	0.017***	0.015***	0.015***	0.012*	0.049***	0.017***	0.024***	0.024**
	[0.013, 0.034]	[0.009, 0.025]	[0.006, 0.023]	[0.007, 0.023]	[0.000, 0.024]	[0.028, 0.069]	[0.005, 0.029]	[0.007, 0.042]	[0.005, 0.044]
2001-2010	0.031***	0.042***	0.025***	0.039***	0.013*	0.049***	0.027***	0.036***	0.040***
	[0.020, 0.041]	[0.030, 0.053]	[0.013, 0.038]	[0.028, 0.050]	[-0.001, 0.028]	[0.028, 0.071]	[0.012, 0.041]	[0.018, 0.054]	[0.019, 0.060]
2011-2018	0.047***	0.063***	0.040***	0.064***	0.044***	0.083***	0.050***	0.061***	0.071***
	[0.029, 0.064]	[0.044, 0.083]	[0.025, 0.054]	[0.046, 0.081]	[0.024, 0.064]	[0.058, 0.108]	[0.032, 0.068]	[0.036, 0.086]	[0.044, 0.097]
1996-2018	0.035***	0.044***	0.028***	0.043***	0.024***	0.061***	0.033***	0.044***	0.048***
	[0.021, 0.049]	[0.029, 0.060]	[0.017, 0.040]	[0.031, 0.055]	[0.008, 0.040]	[0.042, 0.081]	[0.018, 0.048]	[0.024, 0.063]	[0.025, 0.070]
Estimator	OLS	Lasso	OLS	Lasso	OLS	OLS	OLS	OLS	OLS

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.5. The first step is to residualize outcomes using a regression of the outcome on state-sex and time-sex fixed effects and time-varying covariates. The parameters on the covariates are permitted to vary by treatment status, sex, and by time. These parameters are estimated using pre-period data for all states and post-period data for non-triplicate states. When the final row says "Lasso" then rigorous square root lasso with post-estimation OLS is used to predict outcomes. The estimated parameters are used to residualize all observations in the data. The presented estimates are the weighted averages of these residuals for the triplicate states for the designated time periods. Baseline covariates include share (of ages 25-54 by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. For columns with "More Demographic Controls," then share Black and non-Hispanic, 4 education shares, and 3 age shares are also included. The "Add Population" columns use this broader set of controls plus the log of population size as predictors. Policy controls include PDMP, electronic PDMP, medical marijuana laws, operational and legal medical marijuana dispensaries, the state EITC rate (as a percentage of the federal rate), and log of the minimum wage. The "No ASEC" column excludes the ASEC samples in each year. The "ASEC only" column only uses the ASEC samples. The final column logs the outcomes. Regressions and averages are weighted by CPS sample weights.

Appendix A

Appendix Figures

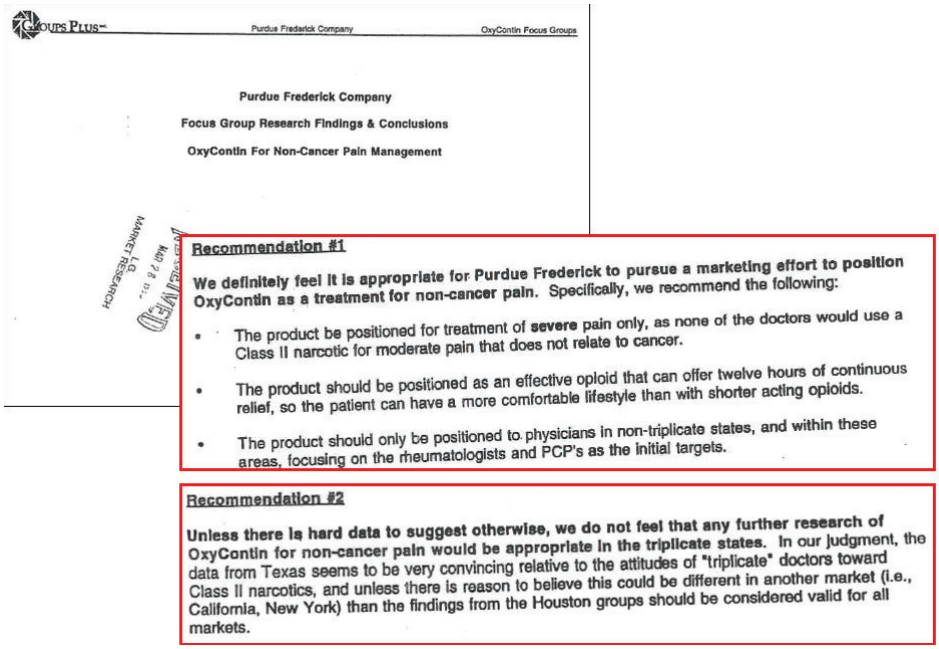
Figure A1: Overdose Death Trends



B: Overdose Deaths

Notes: I use National Vital Statistics System data for 1983-2017. Overdose rates refer to all ages. For 1983-1998, I define drug poisonings as deaths involving underlying cause of death ICD-9 codes E850-E858, E950.0-E950.5, E962.0, or E980.0-E980.5 (see Table 2 of `https://www.cdc.gov/drugoverdose/pdf/pdo_guide_to_icd-9-cm_and_icd-10_codes-a.pdf`, last accessed November 29, 2018.). For opioid-related overdoses, I use deaths involving E850.0, E850.1, E850.2, or N965.0 (Alexander et al., 2018; Green et al., 2017). For the 1999-2017 data, I code deaths as drug overdoses using the ICD-10 external cause of injury codes X40-X44, X60-64, X85, or Y10-Y14 (Warner et al., 2011). I use drug identification codes to specify opioid-related overdoses: T40.0-T40.4 and T40.6. Linking opioid overdoses across ICD-9 and ICD-10 codes in this manner is recommended in Table 3 of `https://www.cdc.gov/drugoverdose/pdf/pdo_guide_to_icd-9-cm_and_icd-10_codes-a.pdf`.

Figure A2: Example of Purdue Pharma Focus Group Recommendations



OUPS PLUS℠ Purdue Fredarick Company OxyContin Focus Groups

Purdue Frederick Company

Focus Group Research Findings & Conclusions

OxyContin For Non-Cancer Pain Management

Recommendation #1

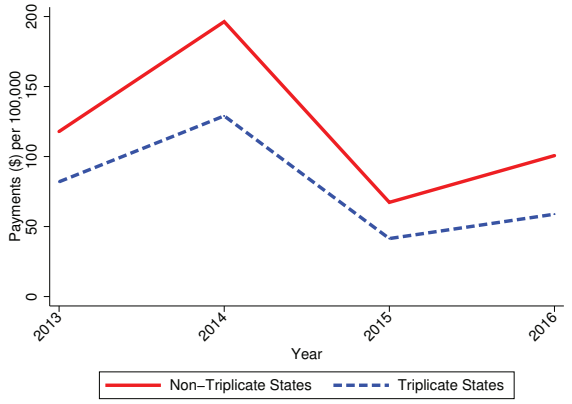
We definitely feel it is appropriate for Purdue Frederick to pursue a marketing effort to position OxyContin as a treatment for non-cancer pain. Specifically, we recommend the following:

- The product be positioned for treatment of **severe** pain only, as none of the doctors would use a Class II narcotic for moderate pain that does not relate to cancer.
- The product should be positioned as an effective opioid that can offer twelve hours of continuous relief, so the patient can have a more comfortable lifestyle than with shorter acting opioids.
- The product should only be positioned to physicians in non-triplicate states, and within these areas, focusing on the rheumatologists and PCP's as the initial targets.

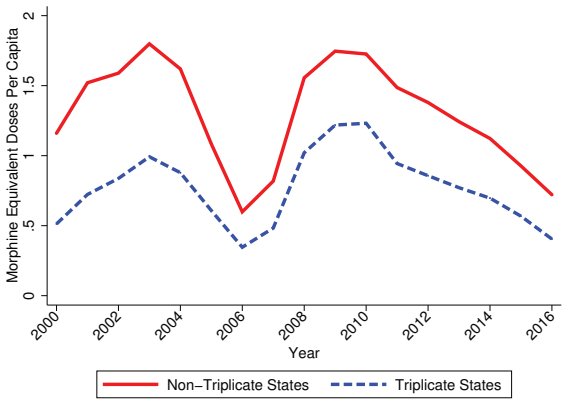
Recommendation #2

Unless there is hard data to suggest otherwise, we do not feel that any further research of OxyContin for non-cancer pain would be appropriate in the triplicate states. In our judgment, the data from Texas seems to be very convincing relative to the attitudes of "triplicate" doctors toward Class II narcotics, and unless there is reason to believe this could be different in another market (i.e., California, New York) than the findings from the Houston groups should be considered valid for all markets.

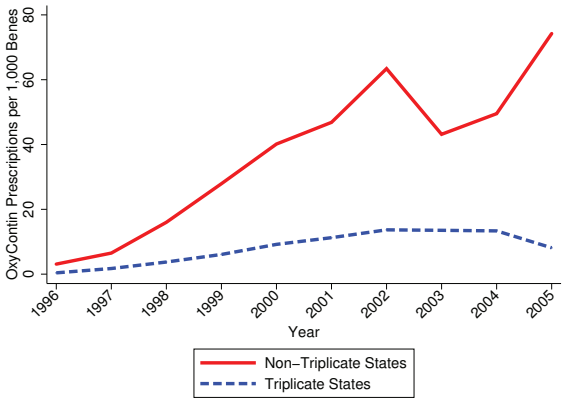
Figure A3: OxyContin Distribution and Prescriptions by Triplicate State Status



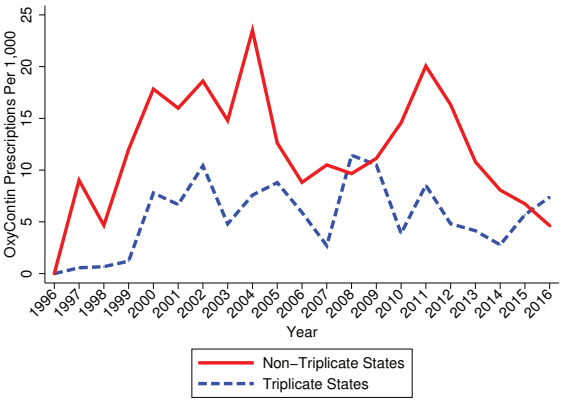
A: OxyContin Promotional Payments



B: OxyContin Distribution (ARCOS)



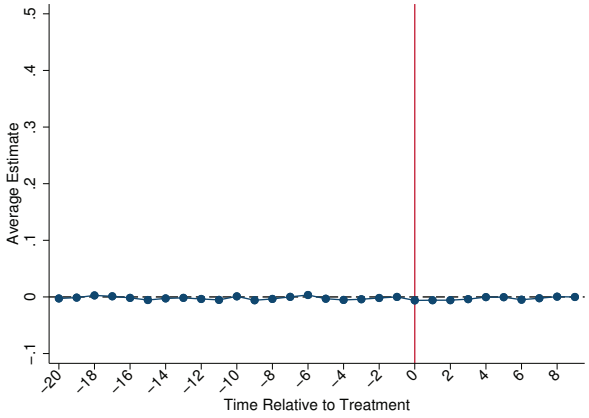
C: OxyContin Prescriptions (Medicaid)



D: OxyContin Prescriptions (MEPS)

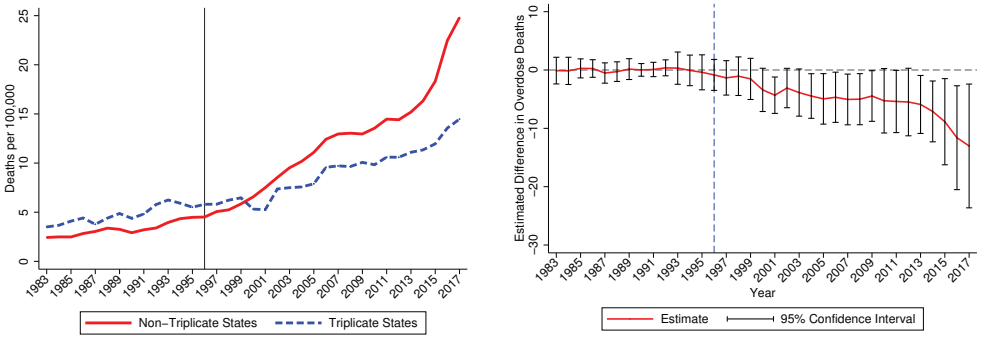
Notes: In Panel A, I use CMS Open Payments Data to calculate total payments and gifts made to physicians regarding OxyContin for the available years. I scale this measure by population. The outcomes correspond to August 2013 – December 2016. Because the 2013 data only cover a partial year, I annualize the rate in that year. In Panel B, I use ARCOS data and construct morphine equivalent doses per capita. OxyContin data are only available for 2000-2016. In Panel C, I report the number of prescriptions per 1,000 beneficiaries from the Medicaid SDUD. I end this time series in 2005 due to the introduction of Medicare Part D. In Panel D, I report the number of prescriptions per 1,000 people in the MEPS. I use the MEPS survey weights.

Figure A4: Simulation Results: Traditional Event Study with No Covariates



Notes: These are the mean bias results from 10,000 simulations detailed in Section 3.1. Here, I present the results when no time-varying covariates are included in the event study estimation.

Figure A5: Overdose Death Results

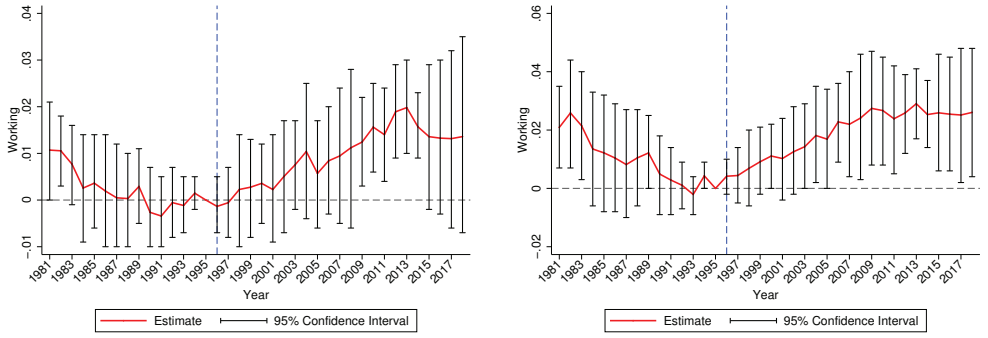


A: Non-Triplicate and Triplicate Trends

B: Adjusted Event Study (Triplicate Minus Non-Triplicate)

Notes: Panel A shows *annual* overdose deaths per 100,000 for triplicate and non-triplicate states for 1983-2017 from the NVSS. See Figure A1 for codes used to determine overdoses in the NVSS data. Panel B provides adjusted event study estimates in which the outcome is *monthly* overdose deaths per 100,000. However, I multiple the estimates (and confidence intervals) by 12 to annualize the results. I residualize the outcome by regressing it on state indicators, month indicators, covariates interacted with treament status, and covariates interacted with year indicators using non-triplicate states and pre-1996 data (i.e., "untreated" observations), weighted by population. I used lasso to select the covariates interacted with year indicators. I report the weighted means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper. Relative to Alpert et al. (2021a), the results are flipped since I consider the triplicate states "treated" here (for reasons discussed in the text).

Figure A6: Traditional Event Study Estimates – No Covariates

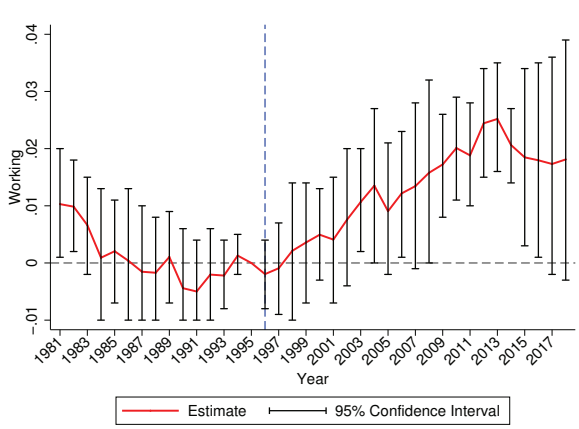


A: Labor Force Participation B: Working

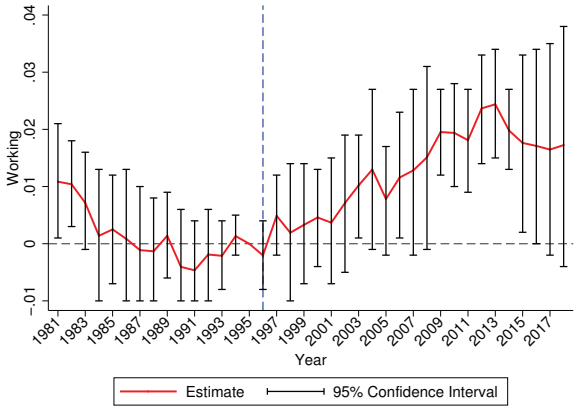
Notes: The outcome is the labor force participation rate or the share working (by month and sex). I regress the outcome on state dummies, time dummies, and year-relative-to-adoption indicators. Regressions are weighted by CPS sample weights. The parameters associated with the covariates are permitted to vary by sex. 95% confidence intervals are also presented and estimated using the procedure discussed in the paper.

Figure A7: Traditional Event Study Results with Covariates

Labor Force Participation Rate

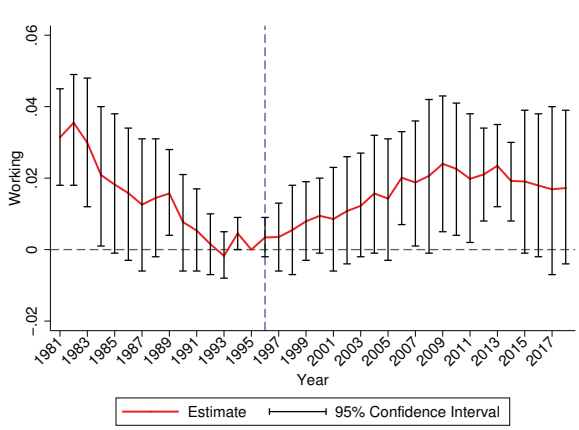


A: Baseline Controls

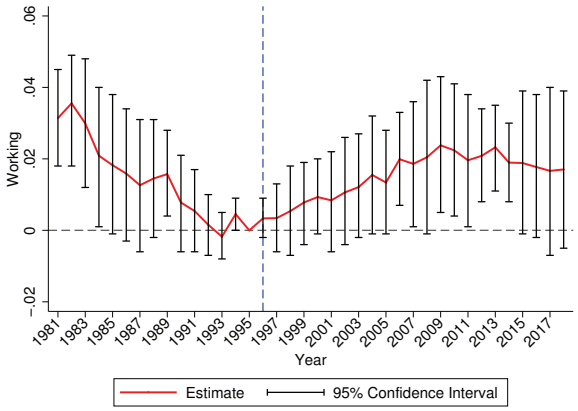


B: Lasso Selects Controls Interacted with Year Dummies

Share Working



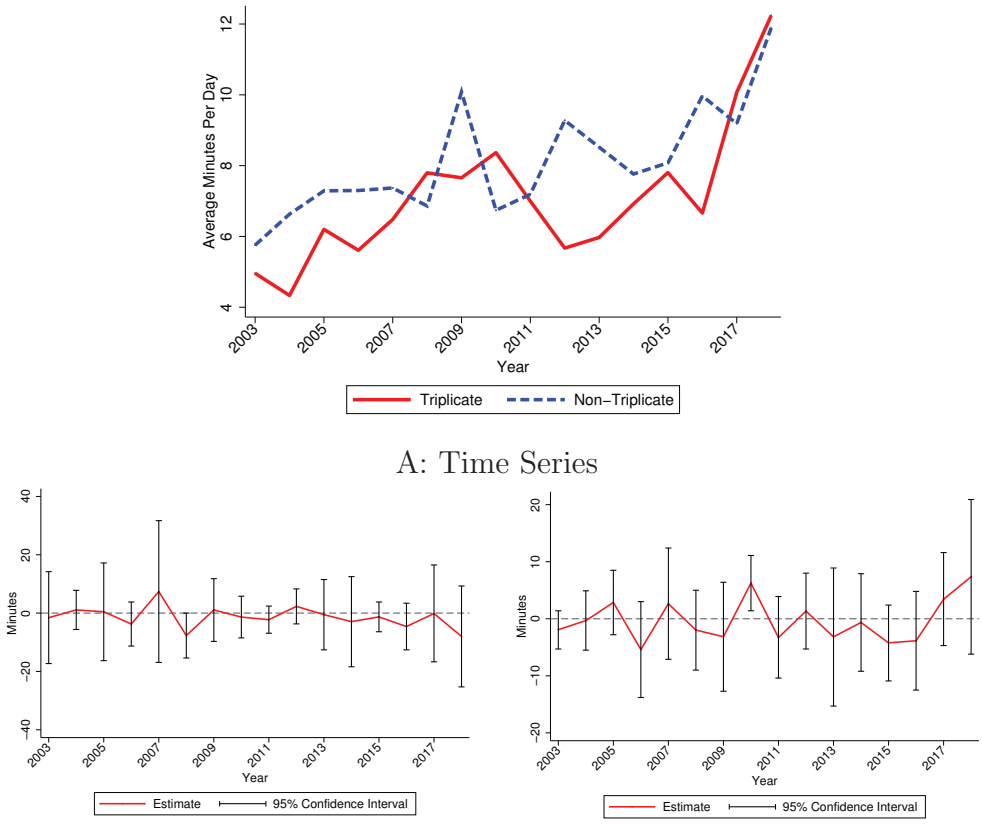
C: Baseline Controls



D: Lasso Selects Controls Interacted with Year Dummies

Notes: The outcome is the labor force participation rate (Panels A-B) or the share working (Panels C-D) by month, state, and sex. I regress the outcome on state-sex dummies, time-sex dummies, and time-varying covariates. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. Regressions are weighted by CPS sample weights. The parameters associated with the covariates are permitted to vary by sex and Triplicate status. In Panels B and D, lasso is used to select covariates with time-varying parameters. 95% confidence intervals are presented using Ferman and Pinto (2019).

Figure A8: Number of Minutes Playing Games



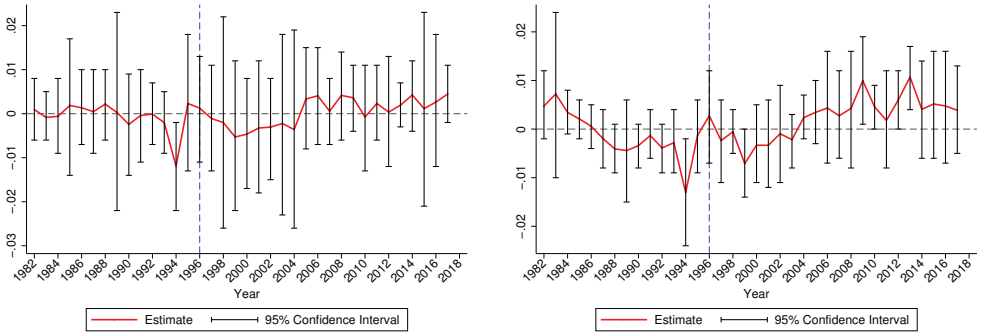
A: Time Series

B: Adjusted Differences

C: Lasso Estimates

Notes: In Panel A, I graph the (unadjusted) average number of minutes "playing games" per day by triplicate status in the ATUS for ages 25-54. In Panels B and C, the outcome is the average number of minutes playing games by year and sex. I residualize the outcome using year-sex interactions as well as covariates which are permitted to have different effects in each year (weighted by population). These covariates are the share white and non-Hispanic, share Hispanic, share with some college, and share ages 45-54. In Panel B, I residualize using all of the covariates interacted with year indicators. In Panel C, I use lasso regression to select the covariates. Note that state indicators are not included since I do not have pre-1996 data for these outcomes. I report the weighted means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper. The estimates refer to adjusted cross-sectional differences.

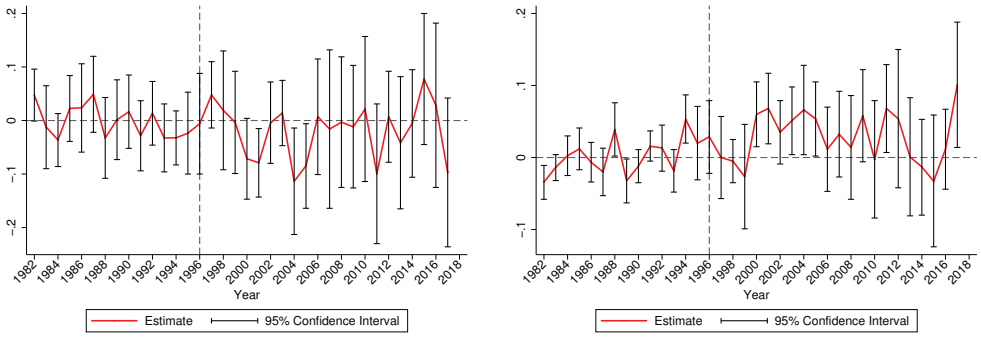
Figure A9: Migration Into and Out of State



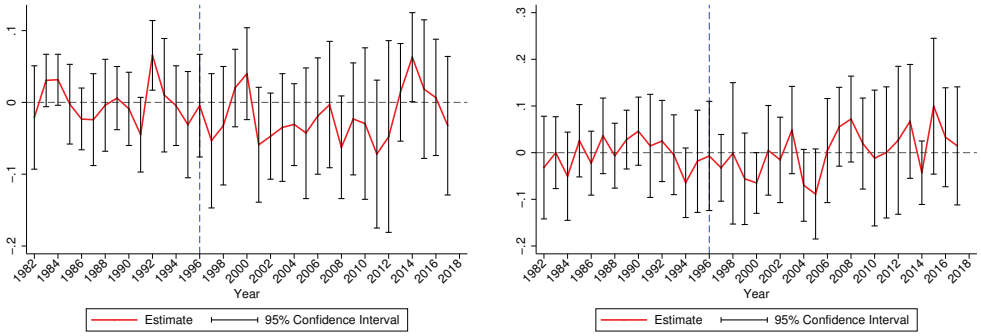
A: Out-Migration B: In-Migration

Notes: The outcome is the share of individuals migrating (or immigrating) into or out of the state (the denominator is the total population in the month-gender cell) by gender. I residualize each outcome by regressing it on state-gender and time-gender indicators using non-triplicate states and pre-1996 data (i.e., "untreated" observations), weighted by population. I report the weighted means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A10: Characteristics of Individuals Migrating Into State



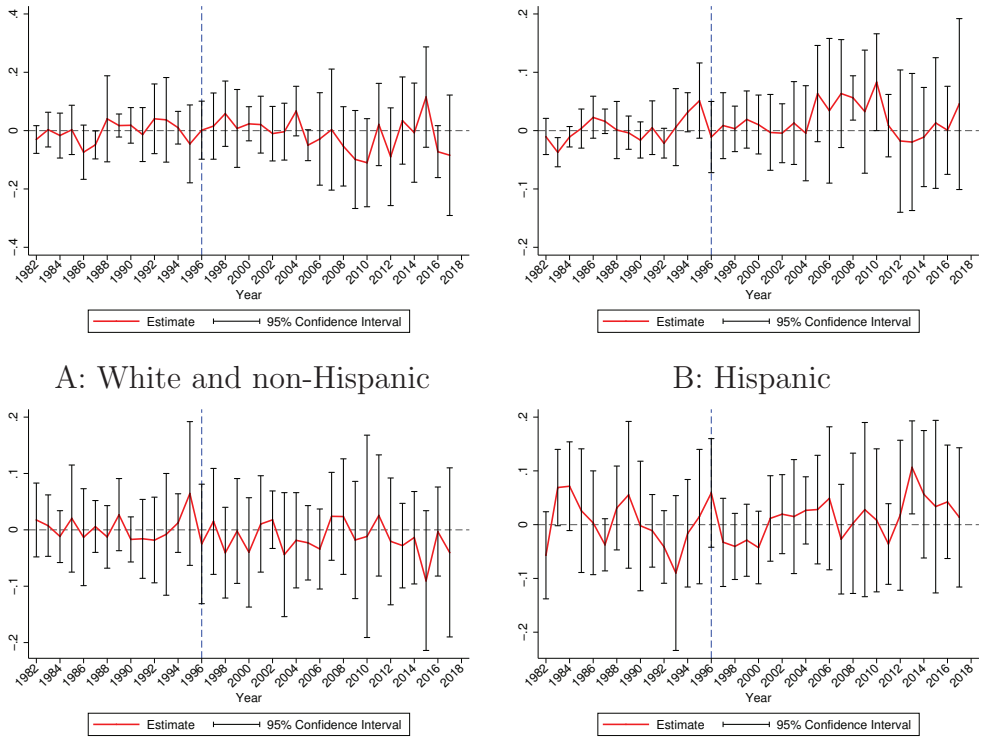
A: White and non-Hispanic B: Hispanic



C: Ages 45-54 D: High Education

Notes: The outcome is the share of individuals migrating into the state with the listed characteristic by month-gender. I residualize each outcome by regressing it on state-gender and time-gender indicators using non-triplicate states and pre-1996 data (i.e., "untreated" observations), weighted by population. I report the weighted means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A11: Characteristics of Individuals Migrating Out of State



A: White and non-Hispanic

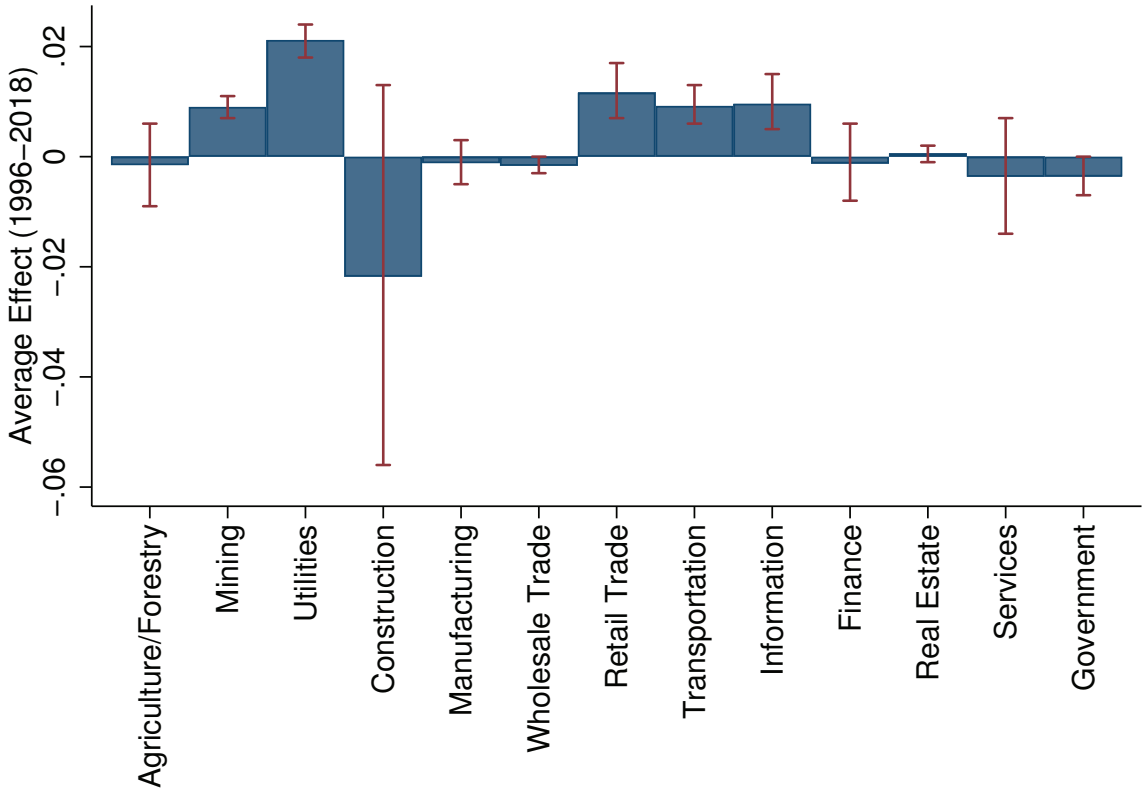
B: Hispanic

C: Ages 45-54

D: High Education

Notes: The outcome is the share of individuals migrating out of the state with the listed characteristic by month and gender. I residualize each outcome by regressing it on state-gender and time-gender indicators using non-triplicate states and pre-1996 data (i.e., "untreated" observations), weighted by population. I report the weighted means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A12: Industry-Specific Results



Notes: The outcome is the share of the 25-54 population working in that industry by sex-month-state. I regress the outcome on state-sex dummies, time-sex dummies, and time-varying covariates using pre-period data and post-treatment data for non-triplicate states. I residualize the outcome and then present annual (weighted by population) averages of these residuals for the triplicate states. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. Regressions are weighted by CPS sample weights. The parameters associated with the covariates are permitted to vary by treatment-sex and year-sex. The year-sex interactions are selected by lasso. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Appendix Tables

Table A1: Traditional Difference-in-Differences Estimates

	A: Labor Force Participation		
Triplicate ×	(1)	(2)	(3)
1996-2000	-0.001	0.008***	0.001
	[-0.004, 0.002]	[0.006, 0.011]	[-0.002, 0.003]
2001-2010	0.007***	0.011***	0.007***
	[0.003, 0.011]	[0.008, 0.015]	[0.003, 0.011]
2011-2018	0.016***	0.020***	0.016***
	[0.011, 0.021]	[0.017, 0.024]	[0.011, 0.021]
1996-2018	-0.006***	0.005***	-0.005**
	[-0.010, -0.002]	[0.002, 0.009]	[-0.008, -0.001]
Covariates	Baseline	Vary by Time	Selected by Lasso
	B: Working		
Triplicate ×	(4)	(5)	(6)
1996-2000	0.002	0.010***	0.003
	[-0.002, 0.006]	[0.006, 0.014]	[-0.001, 0.008]
2001-2010	0.015***	0.016***	0.015***
	[0.010, 0.020]	[0.011, 0.020]	[0.010, 0.020]
2011-2018	0.026***	0.029***	0.026***
	[0.020, 0.032]	[0.024, 0.033]	[0.020, 0.032]
1996-2018	-0.005*	0.006**	-0.004
	[-0.010, 0.000]	[0.002, 0.010]	[-0.008, 0.001]
Covariates	Baseline	Vary by Time	Selected by Lasso

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated using inference procedure discussed in text. Each regression includes state-sex and time-sex fixed effects, time-varying covariates, and a Triplicate indicator interacted with three post dummies. The "1996-2018" result is a result from a separate regression which includes a Triplicate indicator interacted with one post dummy. Covariates include share (of age 25-54 population by sex) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. These are permitted to have different effects by sex and triplicate status. "Vary by Time" means that the covariates are also interacted with year dummies. In the last column, the covariates with time-varying parameters are selected by lasso. The baseline period is 1981-1995. Regressions are weighted by CPS sample weights.

Table A2: Accounting for Labor Demand Shocks

	(1) Main Results	(2) + Bartik Instrument	(3) + Bartik Manufacturing Instrument	(4) + NTR Rate Variables
A: Labor Force Participation				
1996-2000	0.011***	0.012***	0.013***	0.013***
	[0.004, 0.019]	[0.005, 0.018]	[0.007, 0.018]	[0.007, 0.019]
2001-2010	0.021***	0.021***	0.024***	0.026***
	[0.009, 0.032]	[0.013, 0.030]	[0.017, 0.030]	[0.020, 0.033]
2011-2018	0.034***	0.035***	0.038***	0.042***
	[0.022, 0.047]	[0.023, 0.047]	[0.028, 0.049]	[0.032, 0.052]
1996-2018	0.024***	0.024***	0.027***	0.029***
	[0.012, 0.035]	[0.013, 0.035]	[0.017, 0.036]	[0.022, 0.037]
B: Working				
	(5)	(6)	(7)	(8)
1996-2000	0.018***	0.018***	0.020***	0.021***
	[0.006, 0.030]	[0.009, 0.027]	[0.012, 0.028]	[0.009, 0.032]
2001-2010	0.028***	0.029***	0.033***	0.038***
	[0.013, 0.042]	[0.018, 0.039]	[0.026, 0.041]	[0.031, 0.046]
2011-2018	0.051***	0.051***	0.058***	0.064***
	[0.034, 0.069]	[0.035, 0.067]	[0.044, 0.072]	[0.052, 0.076]
1996-2018	0.034***	0.035***	0.039***	0.044***
	[0.019, 0.049]	[0.021, 0.048]	[0.027, 0.052]	[0.036, 0.051]

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.5. The first step is to residualize outcomes using a regression of the outcome on state-sex and time-sex fixed effects and time-varying covariates. The parameters on the covariates are permitted to vary by treatment-sex and by year-sex. I use lasso to select the covariates to interact with year-sex indicators. These parameters are estimated using pre-period data from all states and post-period data for non-triplicate states. Each columns adds a measure of labor demand. The final column uses the Pierce and Schott (2020) measure of exposure to permanent normal trade relations to China. Regressions and averages are weighted by CPS sample weights.

Table A3: Results by Demographics

	(1) Men	(2) Women	(3) Black and non-Hispanic	(4) White and non-Hispanic	(5) Hispanic	(6) No College	(7) Some College
A. Labor Force Participation							
1996-2000	0.002	0.007	0.003	0.003	0.003	0.009**	0.003
	[-0.002, 0.006]	[-0.009, 0.024]	[-0.026, 0.031]	[-0.006, 0.011]	[-0.042, 0.047]	[0.001, 0.017]	[-0.009, 0.015]
2001-2010	0.005	0.030**	0.016	0.01	0.005	0.024***	0.011**
	[-0.001, 0.011]	[0.006, 0.053]	[-0.012, 0.044]	[-0.004, 0.024]	[-0.052, 0.062]	[0.013, 0.035]	[0.002, 0.021]
2011-2018	0.012***	0.049***	0.005	0.020**	0.004	0.040***	0.018***
	[0.004, 0.020]	[0.026, 0.072]	[-0.035, 0.046]	[0.003, 0.036]	[-0.051, 0.060]	[0.017, 0.063]	[0.009, 0.027]
1996-2018	0.007**	0.032***	0.009	0.011	0.005	0.026***	0.012**
	[0.001, 0.013]	[0.011, 0.053]	[-0.018, 0.036]	[-0.002, 0.025]	[-0.050, 0.059]	[0.013, 0.040]	[0.003, 0.022]
B. Working							
1996-2000	0.011***	0.017**	0.013	0.010**	0.005	0.019***	0.008
	[0.003, 0.018]	[0.005, 0.029]	[-0.025, 0.051]	[0.002, 0.018]	[-0.052, 0.062]	[0.007, 0.031]	[-0.004, 0.020]
2001-2010	0.019***	0.051***	0.026*	0.026***	0.016	0.044***	0.023**
	[0.007, 0.031]	[0.033, 0.069]	[-0.003, 0.054]	[0.009, 0.042]	[-0.058, 0.090]	[0.030, 0.058]	[0.007, 0.039]
2011-2018	0.038***	0.076***	0.019	0.038***	0.021	0.067***	0.035***
	[0.028, 0.048]	[0.050, 0.101]	[-0.031, 0.069]	[0.019, 0.058]	[-0.044, 0.087]	[0.052, 0.083]	[0.017, 0.053]
1996-2018	0.024***	0.053***	0.021	0.026***	0.017	0.046***	0.025***
	[0.016, 0.032]	[0.030, 0.076]	[-0.013, 0.054]	[0.010, 0.042]	[-0.051, 0.084]	[0.033, 0.060]	[0.009, 0.040]

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.5. I select on the demographic group listed at the top of each column (for ages 25-54). The first step is to residualize outcomes using a regression of the outcome on state-sex and time-sex fixed effects and time-varying covariates. The parameters on the covariates are permitted to vary by treatment-sex and by time-sex. The covariates with time-sex interactions are selected by lasso regression. Parameters are estimated using pre-period data for all states and post-period data for non-triplicate states. The estimated parameters are used to residualize all observations in the data. The presented estimates are the averages of these residuals for the triplicate states for the designated time periods. Baseline covariates include share (of the relevant population) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. Regressions and averages are weighted by CPS sample weights.

Table A4: Results by Age Group

	(1) Ages 18-64	(2) Ages 25-44	(3) Ages 45-64	(4) Ages 65+
A. Labor Force Participation				
1996-2000	0.005	0.005	0.01	-0.004
	[-0.001, 0.012]	[-0.001, 0.011]	[-0.007, 0.026]	[-0.010, 0.003]
2001-2010	0.013	0.018***	0.014	-0.002
	[-0.002, 0.028]	[0.006, 0.031]	[-0.004, 0.032]	[-0.010, 0.006]
2011-2018	0.022**	0.026***	0.026**	0.006
	[0.004, 0.040]	[0.012, 0.040]	[0.004, 0.047]	[-0.006, 0.019]
1996-2018	0.015*	0.018***	0.018*	0.001
	[0.000, 0.030]	[0.006, 0.030]	[-0.001, 0.037]	[-0.007, 0.009]
B. Working				
1996-2000	0.012***	0.014***	0.014	-0.002
	[0.004, 0.020]	[0.005, 0.023]	[-0.003, 0.032]	[-0.009, 0.004]
2001-2010	0.027***	0.039***	0.023***	-0.001
	[0.010, 0.045]	[0.025, 0.052]	[0.006, 0.040]	[-0.010, 0.009]
2011-2018	0.043***	0.057***	0.036***	0.006
	[0.024, 0.063]	[0.040, 0.074]	[0.016, 0.056]	[-0.009, 0.021]
1996-2018	0.030***	0.040***	0.027***	0.002
	[0.014, 0.047]	[0.026, 0.054]	[0.008, 0.046]	[-0.008, 0.012]

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.5. I select on the demographic group listed at the top of each column. The first step is to residualize outcomes using a regression of the outcome on state-sex and time-sex fixed effects and time-varying covariates. The parameters on the covariates are permitted to vary by treatment status, sex, and by time. These parameters are estimated using pre-period data for all states and post-period data for non-triplicate states. The estimated parameters are used to residualize all observations in the data. The presented estimates are the averages of these residuals for the triplicate states for the designated time periods. Baseline covariates include share (of the relevant population) white and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54 of the 25-54 population. Regressions and averages are weighted by CPS sample weights.

B More Details on Methods

B.1 Formal Conditions for Estimator

Following de Chaisemartin and D'Haultfoeuille (2020), I assume these conditions:

A1 (Balanced Panel) For all $(s,t) \in \{1,\cdots,N\} \times \{1,\cdots,T\}$.

A2 (Policy Adoption) $D_{st} \equiv W_s \times \mathbf{1}(t > T_0)$

A3 (Independence) The vectors $(Y_{st}(0), Y_{st}(1), D_{st}, \mathbf{X}_{st})_{1 \leq t \leq T}$ are mutually independent.

A4 (Common Trends) For $E[\boldsymbol{\gamma}|W_s, t]$,

$$\begin{aligned} &E\left[Y_{st}(0) - \mathbf{X}'_{st}\boldsymbol{\gamma}|W_s = 1, \mathbf{X}_s, t = t_1\right] - E\left[Y_{st}(0) - \mathbf{X}'_{s,t}\boldsymbol{\gamma}|W_s = 1, \mathbf{X}_s, t = t_0\right] \\ &= E\left[Y_{st}(0) - \mathbf{X}'_{st}\boldsymbol{\gamma}|W_s = 0, \mathbf{X}_s, t = t_1\right] - E\left[Y_{st}(0) - \mathbf{X}'_{s,t}\boldsymbol{\gamma}|W_s = 0, \mathbf{X}_s, t = t_0\right] \end{aligned}$$

Condition **A1** enforces that we have a balanced panel. **A2** is a simplifying assumption which requires all treated units to adopt at the same time (with no de-adoption). **A3** assumes independence of the units. **A4** defines the common trends assumption implicit in difference-in-differences designs and permits unit and time heterogeneity. While Section 3 provided a simple model of heterogeneity for illustrative purposes, the necessary conditions for the proposed approach do not require any knowledge of treatment heterogeneity.

A4 relies on $E[\boldsymbol{\gamma}|W_s, t]$ which, as discussed in Section 3.4, does not require a homogeneous paramater for each covariate. In fact, this condition may be more likely to hold given additional flexibility.

B.2 Variance Estimation

Consider the variance of each estimate under the conditions discussed in Section 3.5, including the assumption

$$\epsilon_{st} = \nu_{st} + \sum_{i=1}^{N_{st}} \frac{\omega_{ist}}{M_{st}} \eta_{ist}.$$

There is a unit-year component and an individual component. I assume that the ν_{st} terms are serially-correlated over time with $\nu_s \equiv (\nu_{s1}, \ldots, \nu_{sT})$ i.i.d across s. The η_{ist} terms are i.i.d. Then,

$$\text{Var}(\hat{\beta}_b) = \underbrace{\text{Var}\Big[\sum_{j \in \mathcal{S}_1}\left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\nu_{b(j)t}}{W_{Post,Treated}} - \sum_{t=1}^{T_0} \frac{M_{jt}\nu_{b(j)t}}{W_{Pre,Treated}}\right) - \sum_{j \in \mathcal{S}_0}\left(\sum_{t=T_0+1}^{T} \frac{M_{jt}\nu_{b(j)t}}{W_{Post,Control}} - \sum_{t=1}^{T_0} \frac{M_{jt}\nu_{b(j)t}}{W_{Pre,Control}}\right)\Big]}_{A}$$

$$
\begin{aligned}
&+\left\{\sum_{j\in\mathcal{S}_1}\left[\sum_{t=T_0+1}^{T}\left(\frac{M_{jt}}{W_{Post,Treated}}\right)^2\left(\sum_{i=1}^{N_{b(j)t}}\left(\frac{\omega_{i,b(j),t}}{M_{b(j)t}}\right)^2\right)+\sum_{t=1}^{T_0}\left(\frac{M_{jt}}{W_{Pre,Treated}}\right)^2\left(\sum_{i=1}^{N_{b(j)t}}\left(\frac{\omega_{i,b(j),t}}{M_{b(j)t}}\right)^2\right)\right]\right.\\
&\left.+\sum_{j\in\mathcal{S}_0}\left[\sum_{t=T_0+1}^{T}\left(\frac{M_{jt}}{W_{Post,Control}}\right)^2\left(\sum_{i=1}^{N_{b(j)t}}\left(\frac{\omega_{i,b(j),t}}{M_{b(j)t}}\right)^2\right)+\sum_{t=1}^{T_0}\left(\frac{M_{jt}}{W_{Pre,Control}}\right)^2\left(\sum_{i=1}^{N_{jt}}\left(\frac{\omega_{i,b(j),t}}{M_{b(j)t}}\right)^2\right)\right]\right\}\sigma_\eta^2\\
&+\Delta X_b\Delta X_b'\operatorname{Var}(\hat{\delta})\\
&\equiv A+Bq_b+\Delta X_b\Delta X_b'\operatorname{Var}(\hat{\delta}).
\end{aligned}
$$

This last relationship holds since $\operatorname{Var}\left[\sum_{j\in\mathcal{S}_1}\left(\sum_{t=T_0+1}^{T}\frac{M_{jt}\nu_{b(j)t}}{W_{Post,Treated}}-\sum_{t=1}^{T_0}\frac{M_{jt}\nu_{b(j)t}}{W_{Pre,Treated}}\right)-\sum_{j\in\mathcal{S}_0}\left(\sum_{t=T_0+1}^{T}\frac{M_{jt}\nu_{b(j)t}}{W_{Post,Control}}-\sum_{t=1}^{T_0}\frac{M_{jt}\nu_{b(j)t}}{W_{Pre,Control}}\right)\right]$ is the same across all placebo samples. This equality holds because the weights (on each $\hat{\theta}$) are unchanged.

The q_b term (the term within the $\{\}$ brackets) accounts for heteroscedasticity across placebo samples. I ignore the variance due to estimation of the unit and time fixed effects since both sets of fixed effects cancel out when constructing the overall estimate (or the placebo estimates). The residualization step already produces an estimate of the variance of $\hat{\delta}$. I use the traditional cluster covariance matrix estimator (i.e., "cluster by state") to estimate the variance $\widehat{(\operatorname{Var}(\hat{\delta}))}$. This variance estimate is valid asymptotically given a large number of untreated observations, which should hold in many applications.

Once the variance is estimated, this estimate can be used to scale the estimate. The proposed bootstrap procedure, then, is valid and follows the same process as Ferman and Pinto (2019).

B.3 Differential Levels of Treatment

In the application of this paper, all states are potentially affected by the introduction of OxyContin, but treatment levels vary.[47] I consider the case in which there are treated and untreated units, define as before by W_s, which predicts the *level* of treatment in the following manner:

$$E[D_{st}|W_s,t]\equiv\kappa\mathbf{1}(t>T_0)+(\rho-\kappa)\mathbf{1}(W_s=1,t>T_0),$$

where $\rho>\kappa$. For illustrative purposes, let us assume equation (1) is the true model again. If the researcher assumes a constant value for γ and estimates this relationship using only

[47] I observe useful proxies of the level of treatment by triplicate status.

pre-period data, then

$$E[\hat{\theta}_{st}|W_s = 1, t > T_0] - E[\hat{\theta}_{st}|W_s = 0, t > T_0] = (\rho - \kappa)\beta_0 + \phi_0 \left(\rho E[X_{st}|W_s = 1, t > T_0] - \kappa E[X_{st}|W_s = 0, t > T_0]\right)$$

This represents the causal change in outcomes between treated and untreated units. This metric is useful and standard for what we often estimate in such difference-in-differences designs. In sharp designs, $\rho = 1, \kappa = 0$ so we get $\beta_0 + \phi_0 E[X_{st}|D_{st} = 1]$ as before.

To the extent that there are heterogeneous effects dependent on covariates, then permitting time-specific covariate effects among the $W_s = 0$ group accounts for the impacts of treatment for these units as well as any secular changes in the independent effects of the covariates. This flexibility permits estimation of the counterfactual for the treated units if they had been untreated. Under the equation (1) model, then we estimate

$$E[\hat{\theta}_{st}|W_s = 1, t > T_0] - E[\hat{\theta}_{st}|W_s = 0, t > T_0] = (\rho - \kappa) \left[\beta_0 + \phi_0 E[X_{st}|W_s = 1, t > T_0]\right]. \tag{7}$$

This metric represents the causal change in the outcome due to the additional treatment received by the "more treated" group (i.e., a "treatment on the more treated" estimate). The same assumptions as expressed in Section B.1 are needed (and sufficient) here, where $Y(0)$ should be interpreted as the outcome for $Y(Z_{st} = 0)$ for $Z_{st} = W_s\mathbf{1}(t > T_0)$. However, **A4** takes on a different meaning since this condition will not necessarily hold if there are heterogeneous treatment effects. If those treatment effects vary based on observable characteristics which are systematically different by treatment status, then permitting time-specific heterogeneity in $\boldsymbol{\gamma}_{wt}$ is necessary to recover the counterfactual for the treated units and estimate the "treatment on the more treated." In the case in which all units are treated to varying degrees after T_0, it is necessary to estimate treatment heterogeneity on dimensions that vary by treatment status.

The proposed approach encourages estimating time-specific effects for the covariates in fuzzy designs. As discussed above, this approach requires including a large set of variables in the initial regression. I previously explained concerns about estimating all sources of treatment heterogeneity given a small number of treated units. I assume that the number of untreated units is large so this is less of a concern here. If there are a large number of treated units and small number of untreated units, note that it is appropriate to "flip" what is considered treatment and predict how the outcomes in the less treated units would have evolved given more treatment, identifying the effect of less treatment. In the analysis of this

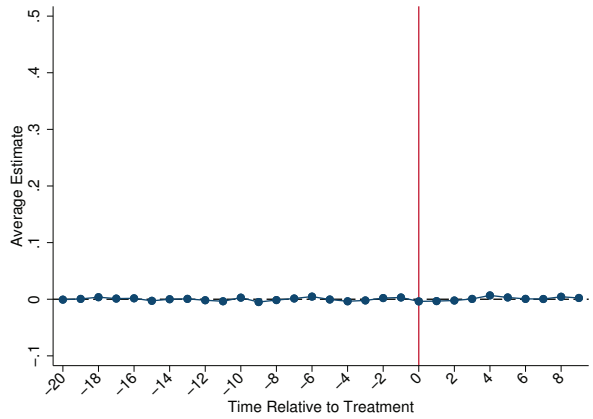
paper, I select the larger group (non-triplicate states) as the control group.

I present simulation results similar to those shown above in Section 3.1 with a slight modification. Define

$$z_{st} = \left\{ \begin{array}{ll} 1, & \text{for } t > 20, s \leq 5 \\ 0, & \text{otherwise} \end{array} \right\}, \quad d_{st} = z_{st} + 1(t > 20).$$

All units are treated in the post-period, but five are more exposed to treatment than the others. The results are provided in Figure B1. In this case, the modified approach first estimates a specification, permitting treatment- and time-specific heterogeneity using only observations in which $t \leq T_0$ *or* $W_s = 0$. The modified approach works well in this context.

Figure B1: Simulation Results: Fuzzy Design



Mean Bias

Notes: These are the results from 10,000 simulations detailed in Section B.3. I present the results using the residualization approach discussed in the paper. The residualization includes unit fixed effects, time fixed effects, and the covariate. The covariate parameter is permitted to vary by treatment status and (additively) by time.

C Trends in Covariates

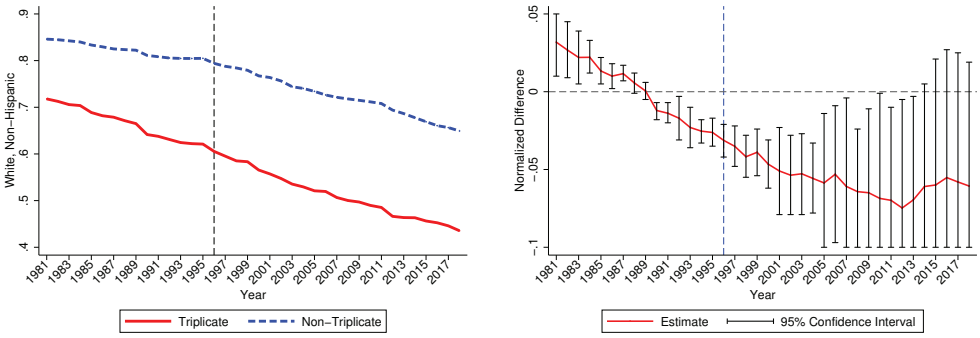
The unconditional labor supply trends in triplicate and non-triplicate states appear different prior to 1996.

In this section, I consider covariates which typically explain labor supply patterns. Given the estimation strategy of this paper, there are benefits to using a small set of covariates for the analysis. I focus on a small set in which I observe differential trends between triplicate and non-triplicate states. The trends are shown in Figure C1. The share that is white and non-Hispanic decreases steadily over time in triplicate states relative to non-triplicate states. This relative reduction is mirrored by a differential rise in the share that is Hispanic. This trend, however, noticeably flattens around 1994-1995 though the difference between the two groups of states still increases through the end of the time period.

The share of the population with at least some college experience declines in triplicate states relative to non-triplicate states. This decline is relatively linear for most of the time period. Finally, I study the share of the 25-54 population which is between the ages of 45-54 since having more of the working-age population at the top end of the age distribution may affect aggregate labor outcomes. I observe a relative decrease in triplicate states, primarily during the 1994-2000 period.

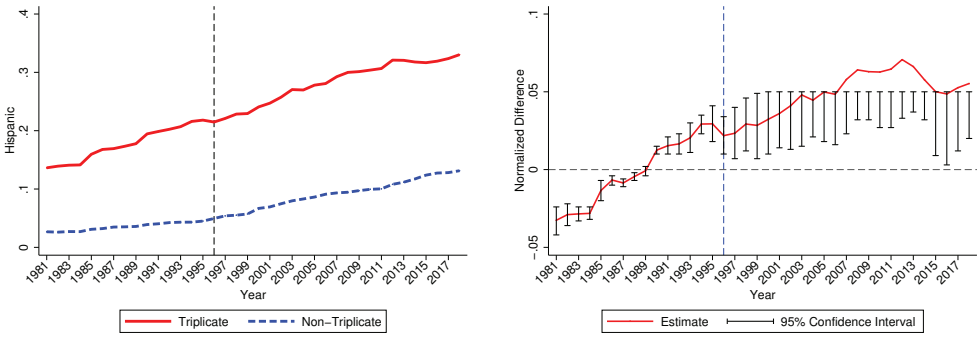
Figure C1: Covariate Trends by Triplicate Status

White, Non-Hispanic



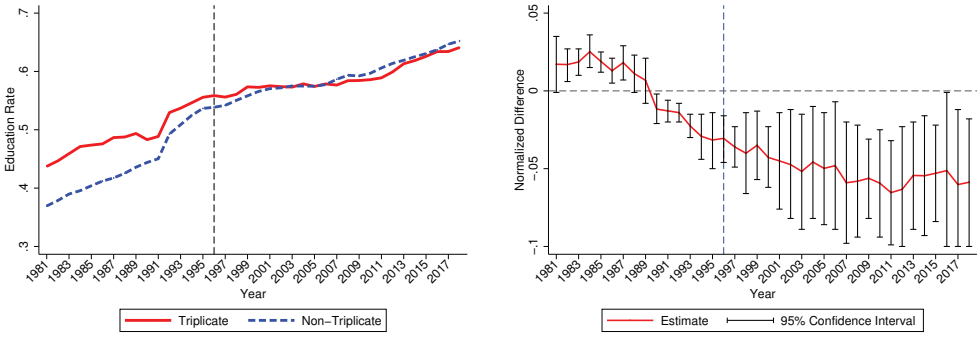
A: Time Series B: Differences

Hispanic



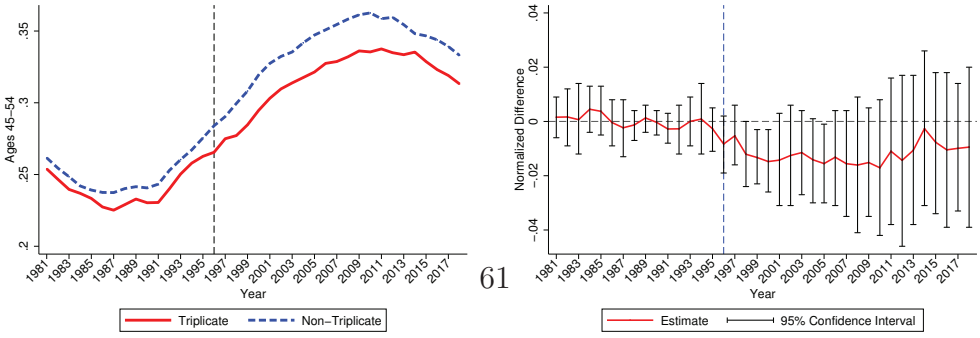
C: Time Series D: Differences

High Education



E: Time Series F: Differences

Ages 45-54



G: Time Series H: Differences

Notes: I graph the covariates over time by triplicate status. I also include normalized annual differences. Differences are normalized to 0 in 1995. 95% confidence intervals are generated using the inference procedure proposed in the paper.