

The Labor Supply Consequences of the Opioid Crisis*

David Powell[†]

November 2022

Abstract

Labor force participation and employment rates have declined in the United States since the late 1990s. An emerging literature considers the role of the opioid crisis on labor outcomes. This paper uses the introduction of OxyContin and geographic variation in its launch to study the long-term labor supply consequences of the opioid crisis. I use an event study framework but show that a standard event study model with covariates can produce biased estimates. I implement a simple modification to this framework. The results suggest that the opioid crisis has played a meaningful role in reducing national labor supply.

Keywords: National Labor Supply Trends, Event Study Designs, Two-way Fixed Effects Models, Substance Use

JEL Classification: J21, I12, C33

*I thank participants at the University of Pennsylvania Center for Health Incentives & Behavioral Economics, the University of Houston Macroeconomics seminar, the Urban Institute, the Center for Health Economics & Policy Studies at San Diego State University, Pontificia Universidad Católica de Chile, RAND, the 2020 APPAM Fall Research Conference, the 2021 American Economics Association Meetings, the 2021 Society of Labor Economists Conference, the 2021 Conference of the American Society of Health Economists, the 2021 Econometric Society North American Summer Meetings, the Michigan Retirement and Disability Research Center, and the RIDGE Health Economics Workshop for helpful discussions and comments. I received especially thoughtful suggestions from Abby Alpert, Miguel Alquezar-Yus, David Beheshti, Breno Braga, Anne Burton, Jessie Coe, Alex Hollingsworth, Jeanne Lafortune, Vegard Nygaard, Isaac Opper, Rosalie Pacula, Evan Peet, Joseph Sabia, Rosanna Smart, and Atheendar Venkataramani. I am thankful for help from Ray Kuntz in accessing the MEPS data at the AHRQ Data Facility. I am grateful for financial support from NIDA (P50DA046351) and CDC (R01CE02999).

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

1 Introduction

Labor force participation and employment rates in the United States have declined since the late 1990s. Figure 1 shows these trends from 1981 to 2018 for the 25-54 age group. The literature has offered varying hypotheses for this decline.¹ Demographic changes are often found to be important determinants of the fall in labor force participation rates (Fernald et al., 2017; Krueger, 2017); however, employment rates declined even within fixed age groups. Declines in labor demand, such as those due to import competition and technology, explain some of the reductions in employment rates (Abraham and Kearney, 2020).

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; Maestas and Sherry, 2020; Beheshti, 2022; Park and Powell, 2021; Cho et al., 2021; Alpert et al., 2022a). The rise of the opioid crisis – see Figure A1 for trends in overdose death rates – occurred as labor force participation rates in the United States dropped. The opioid crisis is a national emergency, affecting the United States on countless dimensions (Maclean et al., 2022), 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.² Krueger (2017) suggests that the growth in opioid access can explain a meaningful share of the decline in labor force participation since 1999. However, Abraham and Kearney (2020) evaluate the role of the opioid crisis in the reduction of employment rates as “unclear” due to concerns about reverse causation in the literature.

In fact, 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, 2019; Venkataramani et al., 2020; Pierce and Schott, 2020; Betz and Jones, 2018; Charles et al., 2019; Currie et al., 2019),³ suggesting that the causal effects of the opioid crisis and economic conditions operate in both directions. The “deaths of despair” hypothesis argues that the opioid crisis is part of a broader trend in deaths related to

¹For some examples, see Aaronson et al. (2012); Hotchkiss and Rios-Avila (2013); Juhn and Potter (2006); Black et al. (2016); Council of Economic Advisers (2014).

²<https://www.cnn.com/2019/07/10/jerome-powell-says-economic-impact-of-opioid-crisis-is-substantial.html>, last accessed April 27, 2020.

³Carpenter et al. (2017) studies prescription pain reliever misuse and economic conditions. Currie and Schwandt (2021) and Maclean et al. (2021) discuss these issues in detail.

weakening socioeconomic and cultural conditions (Case and Deaton, 2015, 2017).⁴

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. The limited set of papers exploring quasi-experimental identification of the impact of the opioid crisis on labor outcomes typically evaluate variation initiated since 2010. 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. I explore the role of the opioid crisis as a major labor supply shock while considering the independent and complementary influences of several labor demand shocks studied in the literature.

Recent 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., 2022b); my focus on OxyContin is due to its pivotal role in the opioid crisis (Kolodny et al., 2015). OxyContin quickly became a blockbuster drug and Purdue Pharma aggressively marketed the use of strong opioids, such as oxycodone products, more broadly.⁵ 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.

I implement a difference-in-differences design, comparing labor outcomes in triplicate states to labor outcomes in non-triplicate states both before and after the introduction of OxyContin, relying primarily on event studies to transparently trace the conditional trajectories of these outcomes over time. 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

⁴Other work concludes that the rise in access to opioids has had independent effects on overdose deaths rates (Ruhm, 2019; Powell et al., 2020; Alpert et al., 2022b).

⁵Purdue 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).

approach, implemented regularly in difference-in-differences designs, is to include controls for these demographic shifts in the specification additively.

This approach can induce bias. The covariates partially fit the treatment heterogeneity in the post-period such that changes in covariates over time 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 designs⁶ with special interest in evaluations of policies with staggered adoption.⁷ In this paper, I highlight that typical implementations of difference-in-differences designs within a regression framework are problematic even without staggered adoption. 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. I also suggest using a principled approach to selecting covariates and modeling their relationships with the outcome variable to predict counterfactual 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.⁸ I show that regression adjustment in event study models biases 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 untreated units and (2) differential shifts in covariates.

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 due to population size differences across units. This method extends the approach introduced in Ferman and Pinto (2019) to account for the variance of the estimated param-

⁶See Kahn-Lang and Lang (2019); Jaeger et al. (2020).

⁷See de Chaisemartin and D’Haultfoeuille (2020); Goodman-Bacon (2021); Callaway and Sant’Anna (2021); Athey and Imbens (2022); Sun and Abraham (2020); Borusyak et al. (2021); Wooldridge (2021); Gardner (2021).

⁸Recent 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.

eters associated with the covariates.

Following Krueger (2017), the main analysis relies on labor supply metrics from the Current Population Survey (CPS) with a focus on ages 25-54, 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 (which are not driven by migration). After adjusting for a small set of demographics, there is little 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, even as illicit opioids begin to dominate the crisis. The estimates imply that triplicate states, due to their reduced exposure to OxyContin’s launch, experienced relative labor force participation growth and employment-to-population growth for the 1996-2018 period. The estimated 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 other policies targeting opioid access and misuse, functional form assumptions, migration, changes in the relationship between the covariates and the outcomes over time, selective mortality, or overfitting. I consider several types of labor demand shocks studied in the literature and find that the results are robust to accounting for these shocks.

Finally, I discuss the implications of the estimates on explaining national trends in labor force participation. I pair this analysis with an examination of labor demand explanations which have previously been found to be important for understanding employment-to-population ratios. I find that these labor demand shocks have had less of a role in altering labor force participation rates over time, implying that much of the decline in labor force participation is unexplained by the literature. An extrapolation exercise using the estimates from this paper finds a substantial role for the opioid crisis for both labor force participation and share working.

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, discusses their implications for understanding national labor supply trends, and considers them in the broader context of the literature studying reasons that employment rates have declined since the 1990s. Section 6 concludes.

2 Background

2.1 Labor Supply Effects of Opioid Access

2.1.1 Literature

Much of the literature on the impact of opioid prescribing (and over-prescribing) on labor supply relies on cross-sectional variation in prescribing behavior or geographic-specific changes in opioid access, assuming that such variation occurs for reasons unrelated to labor market conditions.⁹ It is rare in this literature to leverage policy-driven variation.¹⁰ A recent exception is Beheshti (2022), exploiting the differential geographic impacts of the 2014 rescheduling of hydrocodone. Cho et al. (2021) and Park and Powell (2021) study the reformulation of OxyContin, the removal of one of the most highly-abused opioids, finding sizable reductions in labor force participation as individuals transitioned to illicit markets.¹¹

Shifts in opioid access in more recent years may have different consequences than variation during the beginning of the opioid crisis. Beheshti (2022), Park and Powell (2021), and Cho et al. (2021) consider the consequences of reducing access to a substance or a highly-abusable version of a substance after years of high levels of opioid access. The launch of OxyContin in 1996 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 reflects the inherent difficulties in finding variation in exposure to the opioid crisis since its origins.

⁹Thingholm (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 effects of opioid prescribing rates on Danish labor outcomes.

¹⁰A small, related literature considers the impact of prescription drug monitoring programs on labor supply outcomes (Kaestner and Ziedan, 2019).

¹¹The literature has also found notable complementary effects on disability claiming (Beheshti, 2022; Maestas and Sherry, 2020; Park and Powell, 2021; Arteaga and Barone, 2022).

2.1.2 Theoretical Considerations

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.¹² 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, in principle, has an ambiguous impact on labor supply behavior. Opioid use may lead to dependence issues, reducing work capacity and productivity. Alternatively, opioid access may improve pain management and increase labor force participation rates.¹³ There are also potentially large general equilibrium effects involved – for example, the rise in opioid dependence may strain health care systems (Florence et al., 2021), expand illicit markets and increase crime (Sim, 2021), and have broader impacts on the local economy (Ouimet et al., 2020; Langford, 2021; Cornaggia et al., 2022a,b). Thus, labor supply effects are not necessarily directly related to individual misuse rates but may involve sizable spillovers. Magnitudes in the literature suggest substantial general equilibrium effects.¹⁴

2.2 OxyContin’s Launch

To explore the effects of the opioid crisis on labor outcomes, I focus on the role of OxyContin. Complementary 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., 2022b). 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 to be delivered all at once. This

¹²An exception is Alpert et al. (2022a), who are able to study rich 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.

¹³Currie et al. (2019) find positive labor supply effects for younger women.

¹⁴For example, the effect sizes in Park and Powell (2021) are often larger than the OxyContin misuse rate. This disproportionate effect may simply 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 under-reported. Alternatively, the results suggest that the general equilibrium effects of increases in substance use are sizable. Notably, the estimates in Park and Powell (2021) (see page 17 of that paper) are smaller than those found in the literature (Krueger, 2017; Aliprantis et al., 2019; Beheshti, 2022), implying that most of the labor supply literature has found evidence of large general equilibrium impacts. Other papers on the opioid crisis have found more direct evidence of general equilibrium effects (Cornaggia et al., 2022a,b).

property made OxyContin especially easy to abuse (Cicero et al., 2005; Van Zee, 2009).

2.2.1 OxyContin Promotion

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 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 of the triplicate forms (see Figure A2).¹⁵

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).¹⁶

2.2.2 Triplicate Prescription Programs

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

¹⁵“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).

¹⁶These documents mention other promotional strategies as well. For example, Arteaga and Barone (2022) note that Purdue Pharma initially considered targeting areas with high rates of MS Contin (primarily prescribed for cancer-related pain) prescribing and use variation in pre-1996 cancer deaths to predict exposure to OxyContin’s launch. However, this source of variation is potentially problematic if cancer incidence and deaths provide information about work capacity, healthcare quality, and possible changes in work capacity in the future. For these reasons, I rely on the variation induced by triplicate status, which should not have this property given the timing of the adoption of triplicate programs.

consequences of the *initial* targeting of OxyContin.¹⁷

Figure A3 provides evidence that non-triplicate states were more exposed to the introduction of OxyContin in terms of promotional activity for OxyContin,¹⁸ state OxyContin supply,^{19,20} OxyContin Medicaid prescriptions,²¹ and per capita prescriptions in the MEPS for ages 25-54 during 1996-2016.²² Across all measures and time periods, I observe more OxyContin exposure, and there is evidence of spillovers to other type of opioids (see Figure A4). It might be surprising that prescribing regulations in 1996 could have such persistent effects on marketing strategies. A primary strategy of the Purdue Pharma sales force was to call and visit the top OxyContin prescribers, and this behavior continued until recently.²³ Thus, contemporary differences in promotional activities reflect variation in initial targeting interacted with a marketing strategy subsequently targeting high-prescribing areas.

3 Difference-in-Differences with Covariates

Before discussing the main empirical analysis of the paper, I consider the common practice of including covariates linearly in a two-way fixed effects specification to motivate

¹⁷Alpert et al. (2022b) dedicate a significant amount of analysis to isolating the underlying mechanisms explaining the differential post-1996 overdose trends, determining that it is primarily driven by a marketing effect and not lingering 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 is of less interest to this literature.

¹⁸I study promotional payments for OxyContin, obtained from the Open Payments data base for August 2013 to the end of 2016. The Open Payment Database collects and lists data on payments – for research, meals, travel, gifts, or speaking fees – from drug companies to physicians and teaching hospitals.

¹⁹The online ARCOS data are provided by ingredient. I made a FOIA request for data on OxyContin specifically. This request was approved for 2000-2016.

²⁰There is a notable dip in national OxyContin supply in 2005-2006 due to a patent dispute at this time. Alpert et al. (2022b) 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.

²¹I use the SDUD for 1997-2005. 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.

²²I accessed these data in the AHRQ Data Facility due to the necessity for geocoded data. Per capita prescriptions for all ages are shown separately in Figure A4, Panel A.

²³Later years saw increased targeting to high prescribers: “McKinsey recommended doubling down on Purdue Pharma’s strategy of targeting high prescribers for even more sales calls...” (Commonwealth of Massachusetts, 2018).

the use of a modified approach.²⁴ Sant’Anna and Zhao (2020) discuss that this approach assumes no covariate-specific trends in the treatment units relative to the control units (see footnote 4 of that paper). In this paper, I pay special attention to the implications of additive covariates for event studies and the evaluation of pre-trends and post-treatment dynamic effects.²⁵

I begin with a simple, restrictive (which will not be imposed in the modified approach) model for illustrative purposes. I assume that all units adopt at the same time and there are T_0 periods prior to treatment. Define W_s as an indicator equal to 1 if unit s is ever treated, such that the treatment variable is $D_{st} = W_s \times \mathbf{1}(t > T_0)$. The true model is

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

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 assume a standard event study model is estimated:

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

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 $E[\hat{\beta}_t | t \leq T_0] = 0$, $E[\hat{\beta}_t | t > T_0] = \beta_0 + \phi_0 E[X_{st} | D_{st} = 1]$.

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. The omitted variable bias formula tells us that $E[\hat{\delta}] = \delta_0 + \phi_0 \mu$. This bias creates problems because the event study estimates reflect changes in $y - \hat{\delta}X$, which is not the same in expectation as $y - \delta_0 X$ when $\phi_0 \mu \neq 0$.²⁶ 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*

²⁴When 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).

²⁵A large literature discusses “full regression adjustment” in the context of experimental data. See Negi and Wooldridge (2020) for a recent paper and a comprehensive literature review. See Słoczyński (2022) for a related discussion about OLS with controls.

²⁶Technically, μ can be zero (or negative), but we generally expect $\mu > 0$ given the mechanical relationship between DX and X .

event study model in equation (2):

$$\begin{aligned} & \left(E[y_{st}|W_s = 1, t = t_1] - E[y_{st}|W_s = 0, t = t_1] \right) - \left(E[y_{st}|W_s = 1, t = t_0] - E[y_{st}|W_s = 0, t = t_0] \right) \quad (3) \\ & = (\beta_{t_1} - \beta_{t_0}) + \underbrace{\phi_0 \mu \left\{ \left(E[X_{st}|W_s = 1, t_1] - E[X_{st}|W_s = 0, t_1] \right) - \left(E[X_{st}|W_s = 1, t_0] - E[X_{st}|W_s = 0, t_0] \right) \right\}}_{\text{bias term}} \end{aligned}$$

δ is asked to fit the relationship between y and X for both treated and untreated observations. It cannot accomplish both, inducing bias. The bias is equal to the size of the heterogeneous treatment effects (ϕ_0) times the conditional relationship between DX and D , multiplied by the differential trend in the covariate.

3.1 Modified Approach

For the modified approach, no assumptions on treatment heterogeneity are required. 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, then for equation (1),

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

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 of unit*time data with different population sizes. Let \mathcal{S}_1 represent the set of treated units. Define M_{jt} as the population size of unit j at time t . The difference-in-differences estimate is²⁷

$$\hat{\beta} \equiv \sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T \frac{M_{jt} \hat{\theta}_{jt}}{W_{Post, Treated}}, \quad \text{where} \quad W_{Post, Treated} \equiv \sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T M_{jt}. \quad (4)$$

²⁷Alternative 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 nearly-identical results.

I define the event study time-specific estimates comparably:

$$\hat{\beta}_t \equiv \sum_{j \in \mathcal{S}_1} \frac{M_{jt} \hat{\theta}_{jt}}{W_{t,Treated}}, \quad \text{where} \quad W_{t,Treated} \equiv \sum_{j \in \mathcal{S}_1} M_{jt}. \quad (5)$$

More 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. The residualization approach appropriately handles these cases. This residualization method (or “imputation method” as described elsewhere in the literature) has recently been proposed in the staggered adoption difference-in-differences literature (Gardner, 2021; Borusyak et al., 2021; Liu et al., 2021). These papers dedicate little consideration to covariates. This point has also been made more recently for covariates specifically (Caetano et al., 2022).

I provide simulation results in Section B.2. By selecting only on untreated observations, the modified approach is able to account for the independent effects of the covariates in a manner unaffected by treatment heterogeneity.

3.2 Residualization Step

A common trends assumption is required for the residualized outcomes (see Appendix Section B.1) and relies on obtaining an estimate of $E[\delta|W_s, t]$, the relationship between the covariates and the outcome. There are benefits to permitting this relationship to vary by treatment status and time: $\delta_{w(s)t}$. This flexibility is especially important for empirical applications with differential degrees of treatment (see Section 3.4 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 untreated observations in the post-treatment period.²⁸ We can assume that there is an additive time-specific component to the expected value of δ such that $E[\delta|W_s, t] = \psi_{w(s)} + \zeta_t$.²⁹ The untreated units in the post-period identify *changes* to the

²⁸Jaeger et al. (2020) 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 depending on the context.

²⁹As one simplification, it is reasonable to restrict $\zeta_t = 0$ for $t \leq T_0$ given that we are less worried about capturing heterogeneity within the untreated period.

returns to the covariates with an assumption that, in the absence of treatment, the treated units would have experienced the same changes.

The assumption of this approach is that the treated and untreated units did not experience differential shocks to the (untreated) returns to any of the covariates post-treatment, a standard exogeneity 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 regularization/penalization techniques to reduce concerns of over-fitting. More parameterized expressions of ζ_t may also help. I propose permitting a rich and evolving relationship between the covariates and the outcome while leveraging techniques which guard against overfitting.

3.3 Inference

Given a small number of treated units and possible heterogeneous treatment effects, inference in this context can be difficult.³⁰ I use an inference procedure related to the method introduced in Ferman and Pinto (2019) but extended to the proposed estimation approach. The general idea is to “impute” the variance using the relationship between the residuals and population size in untreated units.

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 $M_{st} \equiv \sum_{i=1}^{N_{st}} \omega_{ist}$. The idea is to bootstrap. Each treated unit j is mapped to unit $b(j)$ and a placebo effect is estimated:

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

$\hat{\theta}_{jt}$ is replaced by $\hat{\theta}_{b(j)t}$, but the unit*time-specific weights are held constant. This property is important for permitting estimation of the variance. The variances of $\hat{\beta}$ and $\hat{\beta}_b$ are not necessarily equal such that traditional permutation tests may not be appropriate.

Under the null, $\hat{\theta}_{st} = \epsilon_{st} + (\alpha_s - \hat{\alpha}_s) + (\gamma_t - \hat{\gamma}_t) + \mathbf{X}'_{st} (\boldsymbol{\delta}_{w(s)t} - \hat{\boldsymbol{\delta}}_{w(s)t})$. Define $\epsilon_{st} \equiv \nu_{st} + \sum_{i=1}^{N_{st}} \frac{\omega_{ist}}{M_{st}} \eta_{ist}$. The error term has a unit-time component and an individual component. I assume that the ν_{st} terms are serially-correlated over time with $\nu_s \equiv (\nu_{s1}, \dots, \nu_{sT})$ i.i.d

³⁰Some inference procedures, such as the wild bootstrap, assume homogeneity of the design matrix (Canay et al., 2019), which rules out using these approaches with 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).

across s . The η_{ist} terms are i.i.d. Define

$$\mathbf{X}_b \equiv \sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T \frac{M_{jt} \mathbf{X}_{b(j)t}}{W_{Post,Treated}}, \quad \hat{\alpha}_{b,j} \equiv \sum_{t=T_0+1}^T \frac{M_{jt} \hat{\alpha}_{b(j)}}{W_{Post,Treated}}.$$

The variance $\hat{\beta}_b$ of can be written as $\text{Var}(\hat{\beta}_b) \equiv A + Bq_b + \text{Var}(\mathbf{X}'_b \hat{\boldsymbol{\delta}}_{w(s)t} + \sum_{j \in \mathcal{S}_1} \hat{\alpha}_{b,j} + \hat{\gamma}_t)$.³¹ This result and the definition of q_b are discussed in Appendix Section B.3. q_b is a function of population size and survey weights and is the factor which adjusts for heteroscedasticity. I estimate the parameters A and B in a regression of $\hat{\beta}_b^2$ on a constant and q_b , using only placebo estimates. An estimate of $\text{Var}(\mathbf{X}'_b \hat{\boldsymbol{\delta}}_{w(s)t} + \sum_{j \in \mathcal{S}_1} \hat{\alpha}_{b,j} + \hat{\gamma}_t)$ is also included on the right-hand side of this regression, but its effect is constrained to equal one. I use the traditional cluster covariance matrix estimator (i.e., “cluster by state”) to estimate the variance of this term, under the assumption that these parameters are estimated using large numbers of comparison units. Both A and B are non-negative.³²

Estimating the variance of the aggregate estimate, instead of state-specific estimates (and then aggregating), has the advantage that it is straightforward to adjust for the variance associated with $\hat{\boldsymbol{\delta}}_{w(s),t}$.³³ After estimation of the parameters A and B , I construct estimates of the variance for each $\hat{\beta}_b$, represented by $\widehat{\text{Var}}(\hat{\beta}_b)$. The steps are as follows:

1. Calculate estimate $\hat{\beta}$.
2. Create \mathcal{B} samples and estimates using equation (6). Use only untreated units and sample with replacement.
3. Using all placebo samples, estimate A and B .
4. Rescale estimates (main and placebo) by the predicted variance.
5. Reject H_0 at level α if and only if $\frac{|\hat{\beta} - \beta_0|}{\sqrt{\widehat{\text{Var}}(\hat{\beta})}} > \frac{|\hat{\beta}_b|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_b)}} [1 - \alpha]$, where $\frac{|\hat{\beta}_b|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_b)}} [q]$ denotes the q^{th} quantile of $\frac{|\hat{\beta}_1|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}}, \dots, \frac{|\hat{\beta}_{\mathcal{B}}|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_{\mathcal{B}})}}$

Only untreated units are used to construct placebo values since the null hypothesis

³¹The time effect is unaffected by the bootstrapping since it is the same across units.

³²I 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.

³³The proposed approach also adjusts for the variance associated with the unit fixed effects, though this is straightforward with many inference approaches. The (implied) adjustment for the variance associated with the time fixed effects is unnecessary since all comparisons are made within the same time period.

relates to the average (across units) effect and does not provide information about unit-specific effects.³⁴ Thus, the null hypothesis does not provide information about the counterfactual values of the treated units’ outcomes, and these units cannot be used to generate placebo estimates.

This approach shares many properties with the Ferman and Pinto (2019) method. The primary difference is that Ferman and Pinto (2019) use unit-specific differences (post minus pre), rescale those differences to adjust for heteroscedasticity, and then aggregate to construct placebo estimates. Here, I propose constructing the placebo estimates and then rescaling using the predicted variance of this final estimate. One advantage of this order of implementation is that it is straightforward to adjust for the variance of the estimated parameters associated with the covariates.³⁵ In Section 5.3.6, I provide results using several alternative inference procedures.

3.4 Differential Levels of Treatment

There are many applications in which all units are initially untreated and then later treated with varying intensities (e.g., Kearney and Levine (2015); Andersen et al. (2020); Alpert et al. (2018)).³⁶ 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, permitting time-specific covariate effects among the “less treated” group accounts for the impact of treatment for these units as well as any secular changes in the independent effects of the covariates. This flexibility allows estimation of the counterfactual for the more treated units if they had been less treated. The assumption (not unique to the proposed method), however, is that all sources of heterogeneity are accounted for by including these covariate*time interactions. For most of this paper, I interpret the results as estimates of the causal changes between more and less treated units. Appendix Section B.4 discusses this issue further and provides simulation results for this type of setup.

³⁴Sharp null hypotheses could also be tested. In this case, one could independently re-sample the treated units.

³⁵Ferman and Pinto (2019) point out that this term asymptotes to zero, though I find it is an important adjustment in finite samples.

³⁶See de Chaisemartin and D’Haultfoeuille (2018) for a related discussion of “fuzzy” difference-in-differences.

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, permitting construction of the measures of interest for ages 25-54 as well as other age groups.³⁷ 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, treating the ASEC sample as a separate month (i.e., estimating a separate time effect for each ASEC sample year). I also provide estimates without the ASEC and for an ASEC-only (annual) sample. Using monthly data, despite the high serial correlation, provides additional information about the relationship between the covariates and outcomes. However, annual data produce similar results (shown below in Table 2).

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 (“employment”) in the previous week.³⁸ As a complementary metric, I study hours worked in the previous week across all jobs, available in all months beginning in 1989. 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 – in the ASEC. These individual components are each censored at high values and the censoring points change throughout the time period. I do not adjust for censoring; instead, I also examine a complementary outcome on employee compensation using Bureau of Economic Analysis (BEA) data which is not affected by censoring. Because

³⁷I exclude respondents listed as in the Armed Forces throughout the analysis.

³⁸There 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.

earnings refer to the previous calendar year, I use the 1982-2019 CPS ASEC 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.³⁹ I observe small discontinuities in 1992 in the constructed measure, but they do not appear to impact the main results.

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. I use the CPS weights to construct means by sex, state, and time.

4.2 Empirical Approach

I follow the approach discussed in Section 3.1, modeling untreated outcomes as

$$Y_{sgmt}(0) = \alpha_{sg} + \gamma_{gmt} + \mathbf{X}'_{sgmt} \boldsymbol{\delta}_{gw(s)t} + \epsilon_{sgmt},$$

where $Y_{sgmt}(0)$ represents the “untreated” labor outcome for state s in month m and year t for sex g . All parameters vary by sex, given that it is typical to model the dynamics of labor supply for men and women differently (Keane, 2011). The specification includes state-sex and time (month-year)-sex fixed effects. Annual treatment effect estimates are defined as $\hat{\beta}_t = \frac{1}{W_{t,treated}} \sum_{s \in \mathcal{S}_{1;g;m}} M_{sgmt} (Y_{sgmt} - \hat{Y}_{sgmt}(0))$, using similar notation as before.

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 time-varying component of the parameters are estimated using the non-treated states. It is helpful to have a large set of control units given that the time-varying parameters use only the untreated units. The above specification is, therefore, estimated for $N = 47,398$.⁴⁰

³⁹Less than 12 years, 12+ but no college, some college but less than 5 years, and 5+ years of post-secondary education.

⁴⁰There are 1,326 observations per year since there are 13 “months” (including the ASEC sample), men and women, and 51 states. There are 1,196 non-triplicate observations per year.

I provide results in which $\delta_{gw(s)t}$ varies by sex-triplicate status (where $w(s)$ denotes the triplicate status of state s). I also permit the parameters to vary by sex-year in the treated period, using the non-triplicate states to estimate how these parameters evolve over time.⁴¹ Because I permit $\delta_{gw(s)t}$ to vary over time, it is important to avoid including a large set of covariates in \mathbf{X} . I select a small set of covariates, discussed in the next section, while also exploring the consequences of including a larger set of covariates (Section 5.3). To further address concerns of over-fitting, I often focus on results in which I penalize the complexity of the relationship between \mathbf{X} and the outcome. I use square root lasso (Belloni et al., 2011), while permitting heteroscedastic and clustered errors (Belloni et al., 2012, 2016), to select the variables to include in the model and then post-estimation OLS (Belloni and Chernozhukov, 2013) to estimate the counterfactual outcomes.⁴² When using lasso to predict counterfactuals, I include the covariates interacted with year indicators, but I also allow for permanent level and slope shifts in the relationships beginning at any post-treatment point in time (i.e., I interact each covariate with $\mathbf{1}(t \geq t')$ and $\mathbf{1}(t \geq t')(t - t')$ for $t' \in \{1996, \dots, 2018\}$). These parameterized time-specific variables reduce the number of variables needed to alter the relationships over time.

I estimate “event study” year-specific effects of triplicate status.⁴³ In addition, I present more aggregated estimates: 1996-2000, 2001-2010, and 2011-2018. The three sub-period estimates are chosen to summarize effects for 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). I also present an average treatment effect for the full 1996-2018 period. The baseline period is 1981-1995 for all 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 approach proposed in Section 3.3.

⁴¹I 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.

⁴²This method is implemented using `lassopack` in Stata (Ahrens et al., 2019, 2020). I only penalize the interactions of the covariates with time-related variables.

⁴³These are averages over 130 observations (5 states, men and women, and 13 “months,” including the ASEC).

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 variables. I explore these 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 (the top of the 25-54 age distribution) have differential trends across treatment.

To summarize the evidence, I evaluate the implications on labor supply. I estimate the relationship between these demographic variables and the labor force participation rate (and working rate) prior to 1996. I use the estimated coefficients to construct predicted labor force participation rates for the full sample period. This exercise illustrates the method of this paper. Figure 2 presents the observed differences in labor force participation rates between triplicate and non-triplicate states as well as the predicted differences based only on the selected observable characteristics. There is a steep relative trend in observed labor force participation, but this trend can be well-explained by demographic changes as shown by the predicted differences based only on covariate shifts. This fit does not represent overfitting, as explored in Section 5.3.3, and the post-treatment gaps are robust to allowing the relationships between the covariates and the outcomes to change over time. Notably, I observe large post-treatment differences in labor force participation and working that cannot be explained by trends in observables. These differences will be attributed to treatment in the main analysis.

5 Results

5.1 Mortality Effects

Before proceeding to the labor supply analysis, I re-examine the differential impact of OxyContin’s launch on overdose deaths. Alpert et al. (2022b) find 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.⁴⁴ Here, I replicate the overdose rate analysis of Alpert et al. (2022b) but use the

⁴⁴Mortality results with covariates are included in Table 1 and the appendix of Alpert et al. (2022b). 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.

proposed two-step residualization approach with the covariates proposed in Section 4.3. The results are presented in Figure A5 and are similar to the unadjusted results found in Alpert et al. (2022b). I also provide results specifically for the 25-54 age group (see Panels C and D). While the mortality effects are large, they are small relative to the labor supply effects found below (i.e., mortality is a rare event). I ignore potential selection effects due to systematic mortality given the order of magnitude difference in the effect sizes.

5.2 Labor Supply Effects

I present the main labor supply results in Figure 3. There is no evidence of any pre-existing trends. Beginning soon after OxyContin’s launch, there is almost continuous differential growth in both labor force participation and share working through the end of the sample period. This pattern holds regardless of whether the covariates are permitted to have different effects over time (Panels B and D).

I summarize these results in Table 1. In Column 1, I residualize using only the baseline covariates with no time-specific interactions. Column 2 interacts the covariates with year indicators. Due to concerns about over-fitting, I use lasso in Column 3 to model the time-specific relationships between the covariates and outcomes. The results are similar across all columns, suggesting that the estimates are not driven by changes in the returns to the covariates over time. Focusing on Column 3, I estimate a small (and insignificant) increase in labor force participation for 1996-2000. For the 2011-2018 period, I estimate that triplicate states experienced relative growth in labor force participation of 3.1 percentage points, statistically significant from zero at the 1% level. For the full post-period, I find that triplicate states had an additional 2.0 percentage points of labor force participation, on average, due to its reduced exposure to OxyContin (relative to non-triplicate states). For the share working, I observe even larger estimates, including an average effect of 3.7 percentage points.

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 interval excludes values lower than 1.1 percentage points; I can also reject that the percentage of people working grew by less than 1.6 percentage points. In Section 5.6.1, I discuss that the point estimates are at the lower end of those found in the literature examining the relationship between the opioid crisis and labor supply.

The event study estimates suggest rather continuous differential labor supply growth,

consistent with the continuous differential growth in oxycodone access. The labor supply effects continue even after reformulation in 2010, consistent with findings in Cho et al. (2021) and Park and Powell (2021) about the role of illicit markets in reducing labor supply in states more exposed to OxyContin.

For comparison purposes, I estimate traditional event studies for both labor force participation and share working. 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.⁴⁵ 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.⁴⁶

Comparing Figure A7 with Figure 3 suggests that treatment heterogeneity is substantially biasing the pre-treatment estimates in the traditional event study approach. The difference-in-differences results are especially sensitive to treatment heterogeneity.

Given the pre-trends observed when shifting demographic are not accounted for (see Figure A6), researchers might try to address them by estimating a model with state-specific (or treatment-specific) linear trends or using alternative approaches like synthetic control estimation (Abadie et al., 2010), among other options. Methods which use pre-treatment outcomes to forecast post-treatment outcomes do not necessarily adequately account for post-treatment outcome changes due to covariates. Consider a covariate which has a differential trend prior to treatment in the treated states relative to the untreated states, but this differential trend stops (exogenously) at some point after treatment. The counterfactual outcome may also have a different trend prior to treatment, induced by this covariate, but we would not expect the outcome trend to then continue throughout the post-period. To the extent that observables explain pre-treatment shifts, this information should be used.

⁴⁵95% confidence intervals are generated using the Ferman and Pinto (2019) approach.

⁴⁶The difference-in-differences estimates are especially sensitive to the definition of the pre-period, which should be clear from Figures A6 and A7. 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.

5.3 Robustness Tests

5.3.1 Possible Confounders

I consider possible confounders for the results provided in Section 5.2.⁴⁷ In the first column of Table 2, I use a larger set of covariates. I include three race/ethnicity variables, three education share variables, and three age group shares. 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 the time components of the additional variables' relationships with the outcomes are penalized in the same manner as before. The results in Column 2 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-sex and permitted to have differential effects over time). Controlling for population size has little effect on the estimates.

In Column 4, 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.⁴⁸ 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 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 5. The estimates are larger than the main estimates.

In Column 6, I exclude the ASEC sample and, in Column 7, I use only the ASEC sample. Results are consistent across the two analysis samples. Finally, in Column 8, I

⁴⁷ I build on the Table 1, Column 2 estimates for this exercise. I do not use penalization methods as the baseline because many of these analyses test sensitivity to including additional control variables, which may not be selected by lasso. To avoid this possibility, I do not penalize. The conclusions of this exercise are similar, however, if I use lasso throughout Table 2, not just in the specified columns.

⁴⁸The 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).

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.

5.3.2 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.⁴⁹ Betz and Jones (2018) find that overdose death rates respond to labor demand shocks using a Bartik-style variable (Bartik, 1991). I construct a similar variable, predicting the share working by interacting baseline (1995) state-specific industry shares with national (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 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.

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,⁵⁰ 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.

Finally, I also consider robustness to both the “China shock” (Autor et al., 2013, 2014, 2016; Acemoglu et al., 2016) and the adoption of industrial robots using data from Acemoglu and Restrepo (2020). In both cases, the papers use first-differences for select years. I replicate my analysis for the same states and “post” years used in those papers and present the results in Table A3. I then add the corresponding labor demand shock measure and estimate the counterfactual using 2SLS (given the instruments used in the cited papers). The estimates are generally unaffected when accounting for exposure to the trade shock and industrial robots.

⁴⁹I show the Table 1, Column 2 results here for the same reasons as before (see footnote 47), but using lasso produces similar results throughout this table.

⁵⁰Data found at <https://www.aeaweb.org/articles?id=10.1257/aeri.20180396> and aggregated to the state level.

5.3.3 Modeling Assumptions, Overfitting Concerns, and Pre-Trends

I consider the sensitivity of the main results to some of my modeling assumptions. The results are similar if I do not disaggregate by sex (i.e., observations are defined by state-month and covariates do not vary by sex). See Appendix Figure A8. Results are also similar if I do not use monthly data (see Table 2, Column 7 for one example).

In my main models, I penalized some of the complexity between the relationship between the covariates and outcomes. However, I still included the same baseline covariates in all models to maintain some consistency across models with different outcomes, samples, etc. Alternatively, I can penalize the inclusion of *all* covariates and their interactions with treatment status and time (i.e., only state*sex and time*sex fixed effects are not penalized). The results are similar and shown in Appendix Figure A9.

It is also worth noting that the covariates are not simply tracing out a pre-existing differential linear trend. If I include an unpenalized linear trend by triplicate status to the model estimated in Appendix Figure A9, the results are similar and the exact same predictors are selected. See Appendix Figure A10. If I penalize the linear trend, it is excluded so the results are identical to those provided in Appendix Figure A9.

One concern may be that the covariates are fitting a fixed relationship with the outcome and then extrapolating that for the entire period, potentially forecasting a relative change in labor supply that would not have occurred. However, the models above permit the relationship between the covariates and outcome to change over time. The results are similar when I permit a very rich relationship (Column 2 of Table 1) and when I penalize the richness of this relationship (Column 3). Thus, the results are not mechanically due to assigning fixed coefficients to covariates which themselves have differential post-treatment trends.

Finally, I can disaggregate my data even further and define “cells” by the interaction of education groups, sex, age group (25-44 and 45-54), and race (White, Black, Asian, and other). This approach permits the inclusion of cell*state and cell*time fixed effects, richly accounting for demographic variation across states and over time. I include the same predictors as before to account for the general equilibrium effects of changes in demographics.⁵¹

⁵¹For further clarification, I do *not* define the covariates by the shares for each state*cell*time period since, for example, the fraction of a cell that is White is either 0 or 1 and perfectly predicted by the cell*state fixed effects. Instead, I control for the share White by state*sex*time period, as before, to account for the general equilibrium effects of changes in composition. This applies to the other covariates as well.

The results are similar even with these richer fixed effects. See Appendix Figure A11.

5.3.4 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. 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 regression results reflect adjusted cross-sectional differences.⁵²

Figure A12, 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 and adjusted differences are small and switch signs throughout the sample period. Below, I find that the difference in hours worked per week far surpasses even the largest magnitudes implied here.

5.3.5 Migration

I 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 are different if they are due to systematic differences in migration. Second, the empirical strategy of this paper focuses on the role of covariates, and I find that these covariates are important in explaining differences across treated and untreated states. The assumption is that the covariates are not responding to OxyContin exposure. Migration could undermine this assumption.⁵³

I test this assumption by studying migration and immigration directly since the CPS reports residence changes in the past year (except for 1981 and 1985). I study total in-migration (scaled by state population) (Panel A of Figure A13), total out-migration (scaled by state population) (Panel B of Figure A13), composition of individuals migrating into the state (Figure A14), and composition of individuals migrating out of the state (Figure A15). I find little evidence of meaningful systematic migration on any dimension. For example, there is some evidence of a disproportionate in-migration increase in triplicate states, peaking at

⁵²The residualization step only involves year-sex fixed effects and covariates interacted with year-sex dummies. These parameters are estimated using the non-triplicate states only.

⁵³We would not expect any endogenous changes in the composition of birth cohorts to have much of an impact on age 25-54 labor outcomes for most of this time period.

under 2 percentage points (using the maximum post-treatment value minus the smallest pre-treatment value). However, even if *all* of these additional in-migrants worked, they would not have a meaningful impact on state labor supply metrics since baseline working rates are high. Similarly, there are some differential shocks in the compositions of migrants, but they never follow the same pattern as the labor supply results, suggesting little influence.

5.3.6 Alternative Inference Approaches

In Appendix Section D, I show results using a host of alternative inference procedures. Instead of re-sampling only from untreated units, I use the same procedure but re-sample from all units, including treated units. I also provide confidence intervals in which I do not adjust for heteroscedasticity, implementing more standard permutation tests. I also generate confidence intervals using Ferman and Pinto (2019). The conclusions of this paper are not changed using any of these alternative procedures.

5.4 Other Measures of Labor Supply

In this section, I study alternative measures. First, I use data from the Current Employment Statistics (CES) to measure the number of jobs per person. 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. People may also have more than one job in a year. While there are limitations to the CES and the estimates are not directly comparable to the CPS, it offers a useful complementary, establishment-level metric of employment. The CES estimates are provided in Figure 4, Panel A.⁵⁴ The pattern of the estimates is similar to the CPS results. In Panel B, I provide the corresponding results using BEA data. The BEA builds on the CES while including information about the self-employed.⁵⁵ The results are similar to the CES results.

Next, I study labor supply measures which also incorporate intensive margin decisions (in addition to the extensive margin). In Panel C, I study “hours worked last week,”

⁵⁴My controls are still demographic shares for the 25-54 population, but the CES and BEA results are robust to using shares based on the 16+ population.

⁵⁵The 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.

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 population size 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 those observed for the other labor outcomes in this paper.

5.5 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 based on demographics, summarized in Figure A16. I first present estimates (overall averages for 1996-2018) by sex. The point estimates in Figure A16 for labor force participation and employment are larger for women than men (consistent with findings in Krueger (2017)), suggesting that female labor supply was more responsive to OxyContin exposure, though men also experienced large labor supply effects.

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) when compared 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 on the extensive margin (Keane, 2011).

These results are also consistent with national labor force participation trends. Men were experiencing declines in labor force participation rates before the 1990s, and this trend continued for most of the sample period (see Appendix Figure A17). Women sharply increased their labor force participation rates in the 1980s and early-1990s, countering the decline in male labor force participation, until a plateau beginning in the mid- to late-1990s,

followed by a decline. Movement in female labor force participation has generally explained the national decline since the mid-1990s.⁵⁶ The results of this paper find that women experienced the largest effects from the opioid crisis, consistent with movements in overall national trends.

Male labor supply tends to respond more on the intensive margin (Keane, 2011). To consider this margin, I study hours worked (see Figure A16, Panel C). The results are more similar for men and women for this outcome, consistent with large intensive margin responses from men (since the outcome reflects both extensive and intensive margin responses).

Next, I study effects by race/ethnicity. I observe the largest extensive margin effects for the Black and non-Hispanic population, though large effects are also observed for the White and non-Hispanic population and the 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.

Overall, there is some evidence that demographic groups with higher rates of pain reliever misuse are not necessarily the ones that experienced worse labor supply outcomes due to the opioid crisis. I include pain reliever misuse rates for 2002-2009 by subgroup using data from the National Survey on Drug Use and Health (NSDUH) in Figure A18.⁵⁷ Misuse rates are high across all groups. However, there is not necessarily a relationship between a group having a higher misuse rate and being more impacted by exposure to the opioid crisis in terms of labor supply, consistent with important general equilibrium effects (though other explanations are also possible).

The main analysis centered on the 25-54 population given the focus in the literature on this group and its relatively high rates of labor force attachment. I study age-based heterogeneity in this section and present the overall (1996-2018) effects in Figure A19.⁵⁸ First, I estimate effects for the 16+ population.⁵⁹ The estimated effects are comparable, though smaller, to those observed for the 25-54 population. I also estimate effects for the

⁵⁶If men and women had both continued on their early-1990s trends, then we would not have observed a national decline in labor force participation.

⁵⁷These years are aggregated together as part of the Restricted-use Data Analysis System and, conveniently, represent a large share of the first wave of the opioid crisis, which was dominated by prescription (not illicit) pain relievers.

⁵⁸For the age results, I typically define the control variables by shares for that age group. However, the results are similar when I use the same control variables as the main analyses.

⁵⁹For this analysis, I controlled for 2 age share variables: share ages 45-64 and share ages 65+. However, results are similar using the same controls as for the main analyses.

18-64 population. These are also similar to the 25-54 estimates.

Next, I study more disaggregated age groups: ages 25-44, ages 45-64, and ages 65+. The biggest effects are observed for the 25-44 age group. I also estimate 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).

One motivation for the age 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. However, the effect sizes for the 65+ age group are small in magnitude, 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.

5.6 Discussion

5.6.1 Comparison 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 (2022) 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. Krueger (2017) and Aliprantis et al. (2019) assume changes in local opioid supply are exogenous to economic conditions while Beheshti (2022) exploits the rescheduling of hydrocodone in 2014. As one measure of differential exposure, I follow Alpert et al. (2022b) and use 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 morphine equivalent doses⁶⁰ (MEDs) per capita of OxyContin compared to 0.43 in triplicate states, implying 165% higher exposure

⁶⁰I define a dose as equal to 60 morphine milligram equivalents.

in non-triplicate states.⁶¹

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.⁶² 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, 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 0.16 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 the results suggest abnormally large labor force participation responses to changes in opioid supply.

5.6.2 Extrapolation Exercise

In this section, I consider the implications of the main estimates on explaining national trends in labor supply – specifically, the decline in labor force participation and employment from 1999 to 2015, representing the years of interest in Krueger (2017) and other work in the literature (and representing the beginning and end of the major decline in labor force participation). As before, I proxy for differential exposure to OxyContin’s launch using the 2000 ARCOS data. In 2000, triplicate states had 0.43 MEDs of OxyContin; non-triplicates had 1.14. The labor supply estimates, therefore, reflect the impact of a difference in 0.71 MEDs (and corresponding differences in promotional activity, spillovers, etc.).

Because of the typical concerns about extrapolation, I provide two metrics about the overall impact on national labor supply. First, I consider a counterfactual in which all states

⁶¹“Exposure” refers broadly to early access, promotional activity, and subsequent spillovers to the prescribing of other opioids resulting from initial differences. I assume that supply differences adequately reflect proportional differences on these dimensions.

⁶²Using 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 prior OxyContin exposure measure may implicitly incorporate the resulting spillovers which increased oxycodone use more broadly).

were exposed to OxyContin at the same level as triplicate states.⁶³ These results have the benefit of being “in-sample,” yet they are conservative as they assume that triplicate states were unaffected by the opioid crisis. Second, I provide estimates using a more traditional extrapolation exercise by estimating a counterfactual in which there was no national exposure to OxyContin.⁶⁴ Both extrapolations are shown in Panel A of Table 3.⁶⁵ I report how much labor force participation declined due to the opioid crisis given the listed counterfactual.

The estimates in Column 1 assume homogeneity between how the treated units respond to treatment and how the untreated units would have responded. I estimate that OxyContin’s launch explains a 1.5 percentage reduction in labor force participation if we consider the triplicate states as the baseline. If we are willing to extrapolate to the national rate of OxyContin exposure, I estimate that OxyContin reduced labor force participation by 2.9 percentage points.

As an alternative, I relax the homogeneity assumption typically enforced with such extrapolations and provide “homogeneity-conditional-on-observables” estimates. I estimate different treatment effects based on demographics and then assume that the same demographic groups in non-triplicate states would have responded in the same manner. I define demographic “cells” in the same manner as before (for Figure A11). These estimates are then reweighted by the size of these cells in the untreated units. For example, if a cell is unaffected, and this group composes a larger share of the population in non-triplicate states, then non-triplicate states will be predicted to have less of an overall response (everything else equal). These results are presented in Column 2 and are similar to the results assuming unconditional homogeneity.

Because the prior approach requires estimating a large number of treatment effects, I also present results in which I penalize treatment heterogeneity across cells as a middle ground between assuming homogeneity and permitting unrestricted heterogeneity across cells. These results are presented in Column 3. The preferred extrapolation estimate (Column 3) is that OxyContin induced a 2.9 percentage point reduction in labor force participation. This estimate is large, explaining a large share of the reduction in labor force

⁶³All extrapolation exercises in this paper are partial equilibrium exercises. For example, if all states had been triplicate states, then Purdue Pharma may have targeted more promotional resources to each state than we observe for triplicate states.

⁶⁴I assume triplicate states experienced $\frac{0.43}{0.71}$ of the estimated labor force participation impacts, the triplicate state exposure (0.43 MEDs) scaled by the difference of 0.71 MEDs. Non-triplicate states experienced $\frac{1.14}{0.71}$ of the estimated impacts.

⁶⁵I rely on the event study estimates for 1999 and 2015 in Figure 3, Panel B.

participation over that time period. Of course, there is little reason to believe that labor force participation, in the absence of the opioid crisis and other factors discussed in the literature, would not have increased, especially given the prior upward trajectory in national labor force participation. Similarly, the opioid crisis may have interacted with other factors to reduce labor supply, implying co-responsibility for the national reduction in labor supply.

5.6.3 What Other Factors Can Explain Decline in Labor Force Participation?

The exercise above suggests that the opioid crisis has played a large role in the reduction of labor force participation rates in the US. Other explanations have been provided in the literature, though primarily for employment. In this section, I re-evaluate the role of labor demand shocks found to meaningfully affect national employment rates, referring to insights from Abraham and Kearney (2020). More details about these replication exercises and the subsequent extrapolation can be found in Appendix Section E. I focus on the “China shock” (Autor et al., 2013, 2014, 2016; Acemoglu et al., 2016) and growth in adoption of industrial robots (Acemoglu and Restrepo, 2020; O’Brien et al., 2022).

As discussed in Appendix Section E, these changes in labor demand explain meaningful reductions in employment-to-population ratios but appear to have less to say about changes in labor force participation. Replicating prior studies in the literature but substituting labor force participation as the outcome, I estimate only small effects related to both of these labor demand shocks, accounting in total for a 0.3 percentage point decline in labor force participation. This finding does not contradict the literature and is not necessarily surprising. I present the individual estimates in Panel B of Table 3.

These small effect sizes suggest that interactions with the opioid crisis are unlikely to explain the large estimated impact of the opioid crisis. In Appendix Section E, I test for these interactions explicitly and find little evidence for the hypothesis that labor demand shocks interacted with the opioid crisis to produce declines in labor force participation.

The literature has also suggested a role for demographics. To quantify how demographic shifts have reduced national labor force participation, I estimate the relationship between the main covariates in the paper and labor force participation using pre-1996 data. I then calculate the change in labor force participation rates due only to shifts in the covariates. This exercise finds that demographics explain a 0.4 percentage point decline in labor

force participation.⁶⁶

Other explanations are possible but, overall, the decline in labor force participation has generally not been well-explained by the literature. The opioid crisis appears to provide an explanation for meaningful shifts in both labor force participation and employment. Taken together, the extrapolation exercises in Table 3 suggest that labor force participation would have increased in the absence of the opioid crisis, demographic shifts, and the labor demand shocks explored in Panel B.

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, including after the removal of the original formulation of OxyContin, consistent with prior evidence found in Cho et al. (2021) and Park and Powell (2021). Notably, traditional difference-in-differences estimates 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.

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

The large effects estimated in this paper are consistent with magnitudes observed elsewhere in the literature examining the relationship between the opioid crisis and labor supply outcomes. While other papers have not necessarily relied on “exogenous” shocks to opioid access, the similarity of the results suggests that triplicate status may be driving much of the variation exploited in the literature (e.g., Krueger (2017); Aliprantis et al. (2019)). Moreover, the magnitudes in this paper and those in the literature likely imply

⁶⁶This magnitude is smaller than those generated from similar exercises, which primarily consider changes in age shares, in the literature because I account for education and educational attainment has increased over time.

substantial general equilibrium effects. Opioid access did not just affect the people who were prescribed opioids or the people who misused opioids. Instead, the opioid crisis affected the population more broadly, inducing higher crime (Sim, 2021; Mallatt, 2022), increasing health care costs (Florence et al., 2021), harming local governments' abilities to provide public services and infrastructure (Cornaggia et al., 2022a), hurting firm productivity (Ouimet et al., 2020; Langford, 2021) and innovation (Cornaggia et al., 2022b), and altering the country on countless other dimensions (Maclean et al., 2022).

The estimates in this paper suggest that the opioid crisis has played a large role in the decline in labor force participation and employment in the United States. This conclusion does not rule out that alternative mechanisms have also had meaningful impacts. However, much of the literature on the decline in labor supply in the United States partially explains reductions in employment-to-population ratios and appears to have less power in explaining the drop in labor force participation. The opioid crisis, on the other hand, seems to have played important roles on multiple labor supply dimensions.

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,” *Journal of Economic Literature*, 2020, *58* (3), 585–643.
- Acemoglu, Daron and Pascual Restrepo**, “Robots and jobs: Evidence from US labor markets,” *Journal of Political Economy*, 2020, *128* (6), 2188–2244.
- , **David Autor, David Dorn, Gordon H Hanson, and Brendan Price**, “Import competition and the great US employment sag of the 2000s,” *Journal of Labor Economics*, 2016, *34* (S1), S141–S198.
- Aguiar, Mark, Mark Bilis, Kerwin Kofi Charles, and Erik Hurst**, “Leisure luxuries and the labor supply of young men,” *Journal of Political Economy*, 2021, *129* (2), 337–382.
- 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.
- Alprantis, 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, Steve Schwab, and Benjamin D Ukert**, “Opioid Use and Employment Outcomes: Evidence from the U.S. Military,” Working Paper 30110, National Bureau of Economic Research June 2022.
- Alpert, Abby, William N Evans, Ethan MJ Lieber, and David Powell**, “Origins of the opioid crisis and its enduring impacts,” *The Quarterly Journal of Economics*, 2022, *137* (2), 1139–1179.
- 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.
- Arteaga, Carolina and Victoria Barone**, “A Manufactured Tragedy: The Origins and Deep Ripples of the Opioid Epidemic,” 2022.
- Athey, Susan and Guido W Imbens**, “Design-based analysis in difference-in-differences settings with staggered adoption,” *Journal of Econometrics*, 2022, *226* (1), 62–79.
- Autor, David, David Dorn, and Gordon Hanson**, “When work disappears: Manufacturing decline and the falling marriage market value of young men,” *American Economic Review*:

- Insights*, 2019, 1 (2), 161–78.
- Autor, David H, David Dorn, and Gordon H Hanson**, “The China syndrome: Local labor market effects of import competition in the United States,” *American Economic Review*, 2013, 103 (6), 2121–68.
- , – , and – , “The China shock: Learning from labor-market adjustment to large changes in trade,” *Annual Review of Economics*, 2016, 8, 205–240.
- , – , – , and **Jae Song**, “Trade adjustment: Worker-level evidence,” *The Quarterly Journal of Economics*, 2014, 129 (4), 1799–1860.
- 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,” *Journal of Human Resources*, 2022, pp. 0320–10762R1.
- Belloni, Alexandre and Victor Chernozhukov**, “Least squares after model selection in high-dimensional sparse models,” *Bernoulli*, 2013, 19 (2), 521–547.
- , **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.
- , – , **Christian Hansen, and Damian Kozbur**, “Inference in high-dimensional panel models with an application to gun control,” *Journal of Business & Economic Statistics*, 2016, 34 (4), 590–605.
- 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, Xavier Jaravel, and Jann Spiess**, “Revisiting event study designs: Robust and efficient estimation,” *arXiv preprint arXiv:2108.12419*, 2021.
- Caetano, Carolina, Brantly Callaway, Stroud Payne, and Hugo Sant’Anna Rodrigues**, “Difference in Differences with Time-Varying Covariates,” *arXiv preprint arXiv:2202.02903*, 2022.
- Callaway, Brantly and Pedro HC Sant’Anna**, “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- 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.
- Cornaggia, Kimberly, John Hund, Giang Nguyen, and Zihan Ye**, “Opioid crisis effects on municipal finance,” *The Review of Financial Studies*, 2022, 35 (4), 2019–2066.
- , – , **Kevin Pisciotta, and Zihan Ye**, “Local Public Resources and Innovation: Evidence from the Opioid Crisis,” *Available at SSRN 4104835*, 2022.
- 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,” *The ANNALS of the American Academy of Political and Social Science*, 2021, 695 (1), 276–291.
- , **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,” *Brookings Papers on Economic Activity*, 2017, pp. 1–81.
- 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.
- Florence, Curtis, Feijun Luo, and Ketra Rice**, “The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017,” *Drug and alcohol dependence*, 2021, 218,

108350.

- Gardner, John**, “Two-stage differences in differences,” Technical Report, Working paper 2021.
- Goodman-Bacon, Andrew**, “Difference-in-differences with variation in treatment timing,” *Journal of Econometrics*, 2021, 225 (2), 254–277.
- 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*, 2020, 38 (2), 317–326.
- 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.
- Kolodny, Andrew, David T Courtwright, Catherine S Hwang, Peter Kreiner, John L Eadie, Thomas W Clark, and G Caleb Alexander**, “The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction,” *Annual review of public health*, 2015, 36, 559–574.
- 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.
- Langford, Scott**, “We’re Not in Dreamland Anymore: How Regional Opioid Use Rates Affect Industrial Composition,” *Available at SSRN*, 2021.
- Liu, Licheng, Ye Wang, and Yiqing Xu**, “A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data,” *arXiv preprint arXiv:2107.00856*, 2021.
- Maclean, Johanna Catherine, Justine Mallatt, Christopher J Ruhm, and Kosali I Simon**, “Economic Studies on the Opioid Crisis: Costs, Causes, and Policy Responses,” in “Oxford Research Encyclopedia of Economics and Finance” 2021.
- , – , – , and – , “The Opioid Crisis, Health, Healthcare, and Crime: A Review Of Quasi-Experimental Economic Studies,” Working Paper 29983, NBER April 2022.
- Maestas, Nicole and Tisamarie B Sherry**, “Opioid Treatment for Pain and Work and Disability Outcomes: Evidence from Health Care Providers’ Prescribing Patterns,” 2020.
- Mallatt, Justine**, “Policy-induced substitution to illicit drugs and implications for law enforcement activity,” *American Journal of Health Economics*, 2022, 8 (1), 000–000.
- Negi, Akanksha and Jeffrey M Wooldridge**, “Revisiting regression adjustment in experiments with heterogeneous treatment effects,” *Econometric Reviews*, 2020, pp. 1–31.
- O’Brien, Rourke, Elizabeth F Bair, and Atheendar S Venkataramani**, “Death by robots? Automation and working-age mortality in the United States,” *Demography*, 2022, 59 (2), 607–628.
- Ouimet, Paige, Elena Simintzi, and Kailei Ye**, “The impact of the opioid crisis on firm value and investment,” *Available at SSRN 3338083*, 2020.
- 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, Rosalie Liccardo Pacula, and Erin Taylor**, “How increasing medical access to opioids contributes to the opioid epidemic: Evidence from Medicare Part D,” *Journal of Health Economics*, 2020, p. 102286.
- 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.
- 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.

- Sim, Yongbo**, “The Effect of Opioids on Crime: Evidence from the Introduction Of OxyContin,” 2021.
- 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.
- Słoczyński, Tymon**, “Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights,” *The Review of Economics and Statistics*, 05 2022, 104 (3), 501–509.
- 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.
- Wooldridge, Jeff**, “Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators,” *Available at SSRN 3906345*, 2021.
- Zee, Art Van**, “The promotion and marketing of OxyContin: commercial triumph, public health tragedy,” *American Journal of Public Health*, 2009, 99 (2), 221–227.

Figures

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

National Trends

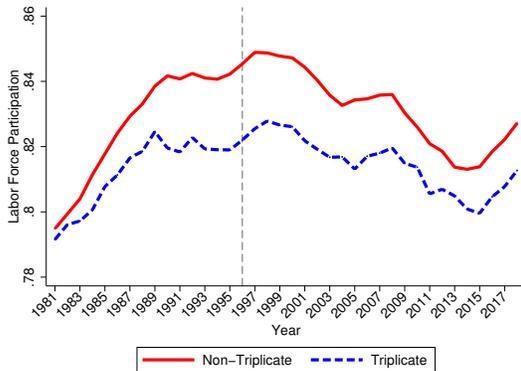


A: Labor Force Participation Rate

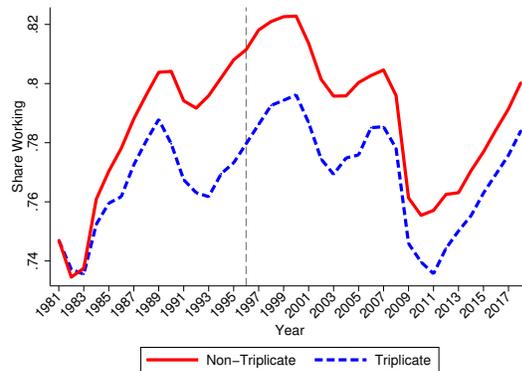


B: Share Working

By Triplicate Status



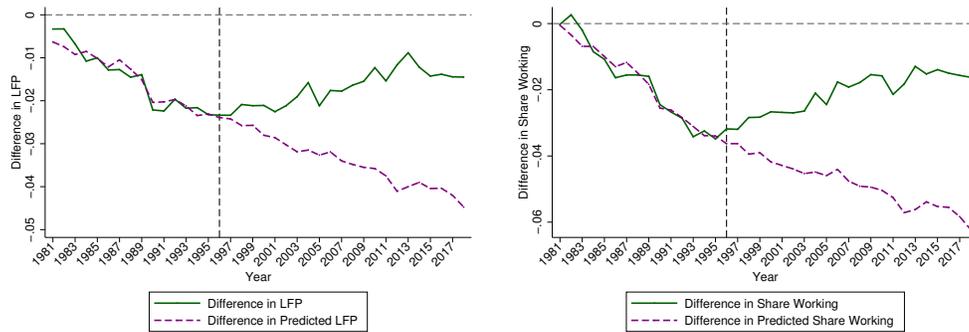
C: Labor Force Participation Rate



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 working. These trends are shown nationally and by triplicate status.

Figure 2: Outcome Differences Between Triplicate and Non-Triplicate States



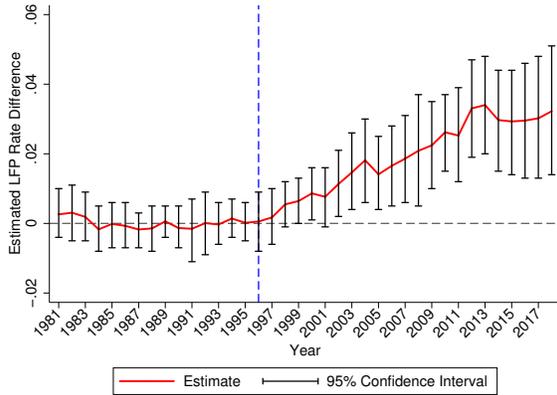
A: Labor Force Participation

B: Working

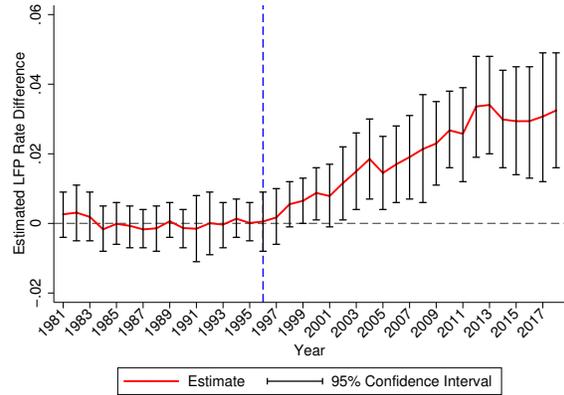
Notes: Each figure plots the observed difference in annual labor force participation rate or share working. It also includes a prediction based only on covariates and fixed effects. Using state*sex*month data, I regress the outcome on state-sex dummies, time-sex dummies, and time-varying covariates using 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 3: Main Event Study Results

Labor Force Participation Rate

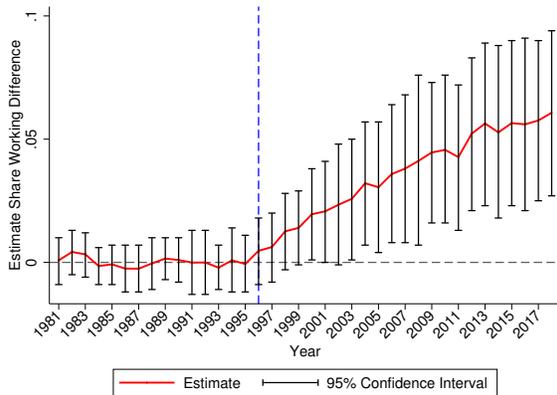


A: Baseline Controls

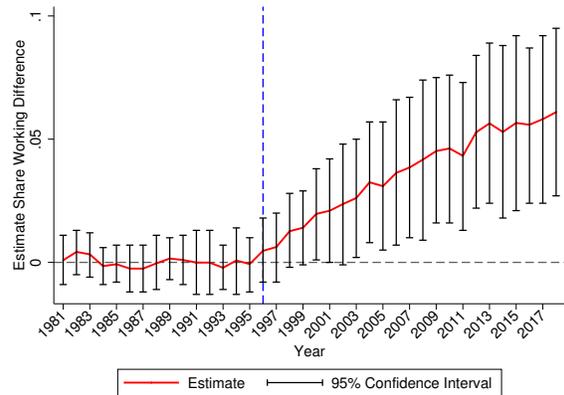


B: Lasso

Share Working



C: Baseline Controls

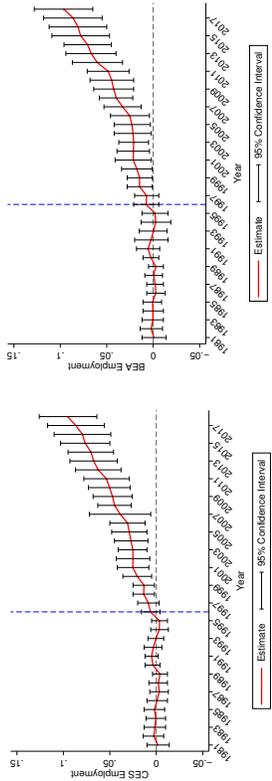


D: Lasso

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 aggregated CPS sample weights) 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 sex-year – lasso is used to select these covariates. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure 4: Event Study Estimates for Other Measures of Labor Supply

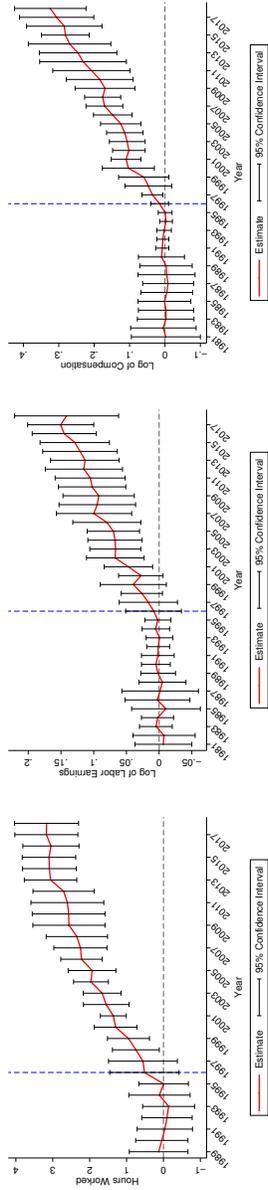
Number of Jobs Per Capita (Ages 16+)



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 (results are similar if covariates are defined as shares of age 16+ population). Regressions are weighted by population size (16+). The parameters associated with the covariates are permitted to vary by Triplicate status. I also let the relationship between the covariates and outcomes vary over time, using lasso to select the time-varying parameters. In Panels C and D, the outcomes are weekly hours worked (available starting in 1989 for all CPS samples) and the log of labor earnings, respectively, in the CPS for ages 25-54. 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: Main Difference-in-Differences Estimates

A: Labor Force Participation			
Triplicate ×	(1)	(2)	(3)
1996-2000	0.005 [-0.002, 0.011]	0.011** [0.002, 0.020]	0.005 [-0.002, 0.011]
2001-2010	0.017*** [0.006, 0.028]	0.021*** [0.006, 0.035]	0.018*** [0.007, 0.028]
2011-2018	0.030*** [0.017, 0.044]	0.034*** [0.017, 0.052]	0.031*** [0.017, 0.044]
1996-2018	0.019*** [0.011, 0.028]	0.024*** [0.012, 0.035]	0.020*** [0.011, 0.028]
Covariates	Baseline	Vary by Time	Selected by Lasso
B: Working			
Triplicate ×	(4)	(5)	(6)
1996-2000	0.011 [-0.003, 0.026]	0.018** [0.001, 0.034]	0.012 [-0.004, 0.027]
2001-2010	0.034** [0.006, 0.062]	0.028* [-0.001, 0.058]	0.034** [0.005, 0.063]
2011-2018	0.054*** [0.027, 0.082]	0.051*** [0.017, 0.085]	0.055*** [0.026, 0.083]
1996-2018	0.037*** [0.015, 0.058]	0.034*** [0.012, 0.056]	0.037*** [0.016, 0.058]
Covariates	Baseline	Vary by Time	Selected by Lasso

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated using the method proposed in Section 3.3. 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 the pre-period for all states and post-treatment period for non-triplicate states. I then report the weighted 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 and triplicate status. “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 (and how they vary over time). Regressions and averages are weighted by CPS sample weights.

Table 2: Robustness Tests

A. Labor Force Participation								
	(1) More Demographic Controls	(2) More Demographic Controls	(3) Add Population	(4) Policy Variables	(5) PDMP States	(6) No ASEC	(7) ASEC Only	(8) Log of Outcome
1996-2000	0.014*** [0.010, 0.018]	0.009*** [0.005, 0.013]	0.008*** [0.004, 0.012]	0.011*** [0.007, 0.014]	0.029*** [0.018, 0.040]	0.011*** [0.007, 0.015]	0.015*** [0.006, 0.024]	0.014*** [0.001, 0.026]
2001-2010	0.022*** [0.018, 0.026]	0.024*** [0.019, 0.028]	0.022*** [0.018, 0.026]	0.018*** [0.012, 0.024]	0.030*** [0.018, 0.042]	0.020*** [0.016, 0.025]	0.021*** [0.008, 0.034]	0.025*** [0.009, 0.042]
2011-2018	0.035*** [0.028, 0.042]	0.039*** [0.031, 0.047]	0.039*** [0.034, 0.044]	0.042*** [0.032, 0.052]	0.046*** [0.031, 0.061]	0.034*** [0.026, 0.042]	0.036*** [0.019, 0.052]	0.042*** [0.019, 0.065]
1996-2018	0.025*** [0.020, 0.030]	0.026*** [0.021, 0.031]	0.025*** [0.021, 0.030]	0.025*** [0.018, 0.032]	0.035*** [0.021, 0.050]	0.023*** [0.019, 0.028]	0.025*** [0.015, 0.036]	0.029*** [0.014, 0.044]

B. Working								
	OLS	Lasso	Lasso	OLS	OLS	OLS	OLS	
1996-2000	0.021*** [0.016, 0.025]	0.016*** [0.011, 0.021]	0.014*** [0.009, 0.020]	0.014*** [0.009, 0.019]	0.043*** [0.029, 0.058]	0.017*** [0.012, 0.021]	0.026*** [0.015, 0.037]	0.023*** [0.002, 0.044]
2001-2010	0.027*** [0.020, 0.033]	0.040*** [0.033, 0.046]	0.037*** [0.031, 0.042]	0.019*** [0.010, 0.027]	0.038*** [0.023, 0.054]	0.027*** [0.020, 0.034]	0.040*** [0.024, 0.055]	0.039* [-0.003, 0.080]
2011-2018	0.046*** [0.039, 0.053]	0.061*** [0.054, 0.068]	0.061*** [0.054, 0.068]	0.054*** [0.044, 0.064]	0.069*** [0.053, 0.086]	0.050*** [0.042, 0.057]	0.063*** [0.042, 0.084]	0.068*** [0.023, 0.112]
1996-2018	0.032*** [0.027, 0.038]	0.043*** [0.035, 0.050]	0.041*** [0.034, 0.048]	0.031*** [0.022, 0.039]	0.051*** [0.031, 0.070]	0.033*** [0.028, 0.038]	0.046*** [0.032, 0.060]	0.046*** [0.017, 0.075]

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.3. 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*sex. 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 select the time*sex component of the relationship between the covariates and the outcome. 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," share Black and non-Hispanic, 3 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 in Column 4 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.

Table 3: Extrapolation Exercise for Labor Force Participation Rate (Ages 25-54) Decline, 1999-2015

Panel A. Extrapolation Exercise (in percentage points)			
Labor Force Participation in 2015 Relative to 1999 due to Opioid Crisis			
	(1)	(2)	(3)
1) Triplicate States as Counterfactual	-1.5	-1.6	-1.5
2) No OxyContin as Counterfactual	-2.9	-2.8	-2.9
Homogeneity/Heterogeneity by Demographic?	Homogeneity	Heterogeneity	Heterogeneity
Heterogeneity in Demographic Treatment Effect Estimates Penalized?	N/A	No	Yes

Panel B. Explanations in Literature for 3.2 Percentage Point Decline

China Shock	-0.069 percentage points
Robots	-0.195 percentage points
Demographics	-0.387 percentage points
Total	-0.651 percentage points

Notes: I use the Figure 3, Panel B estimates for 1999 and 2015. In Panel A, “Triplicate States as Counterfactual” estimates the partial equilibrium effect of the opioid crisis using triplicate states, and their level of OxyContin exposure, as the counterfactual. The Figure 3, Panel B estimates provide estimates of the labor supply reduction incurred by non-triplicate states relative to triplicate states. This approach assumes that triplicate states were not affected by the opioid crisis. “No OxyContin as Counterfactual” estimates how national labor force participation was impacted by OxyContin, extrapolating to no OxyContin exposure. This approach scales the Figure 3, Panel B estimates. Those estimates represent the effects due to a 0.71 MED difference in initial OxyContin exposure. Non-triplicate states are scaled to have experienced $\frac{1.14}{0.71}$ of the estimated labor force participation impacts; triplicate states experienced $\frac{0.43}{0.71}$ of the estimated labor force participation impacts.

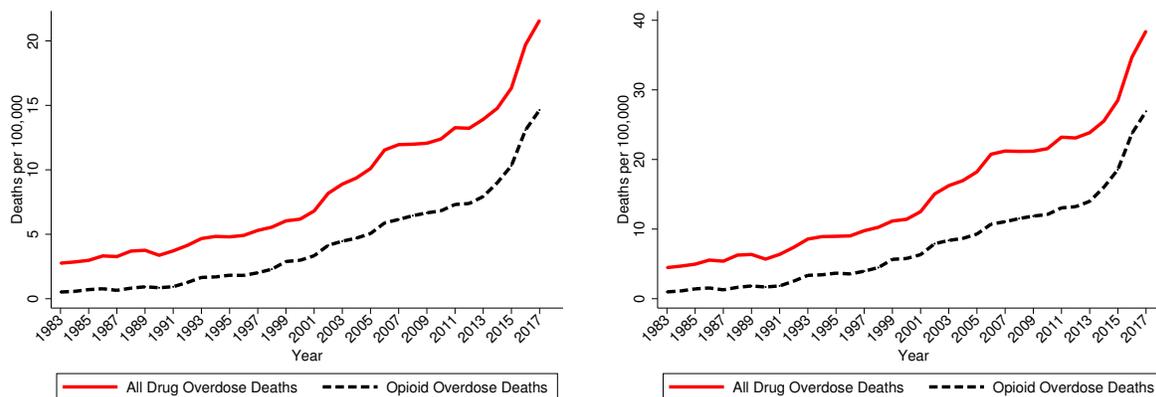
In Columns 2 and 3, I estimate the Figure 3, Panel B event study for every demographic “cell” (defined in text) and assign that effect by cell to non-triplicate states as well as the triplicate states. This approach reweights the cell-specific effects based on the demographic composition of the non-triplicate states. In Column 3, I penalize heterogeneity in the treatment effects when estimating cell-specific effects.

Panel B extrapolates the impacts of demographic changes (see main text for details), the China trade shock (using Autor et al. (2019)), and the increase in industrial robot use (using Acemoglu and Restrepo (2020)). See Appendix E for specifics.

ONLINE APPENDIX: Appendix A

Appendix Figures

Figure A1: Overdose Death Trends



A. Full Population

B. Ages 25-54

Notes: I use National Vital Statistics System data for 1983-2017. Overdose rates refer to all ages in Panel A; ages 25-54 in Panel B. 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).

Figure A2: Example of Purdue Pharma Focus Group Recommendations

GROUPS PLUS Purdue Frederick 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.

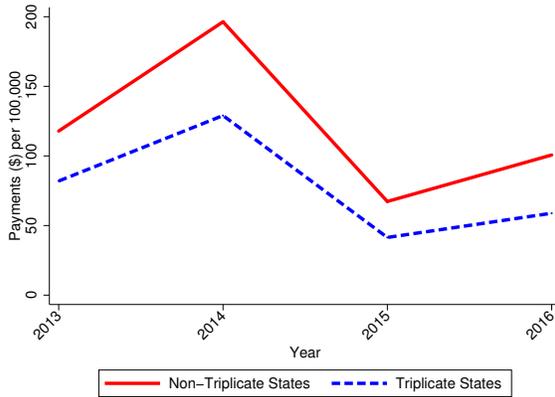
Recommendation #2

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.

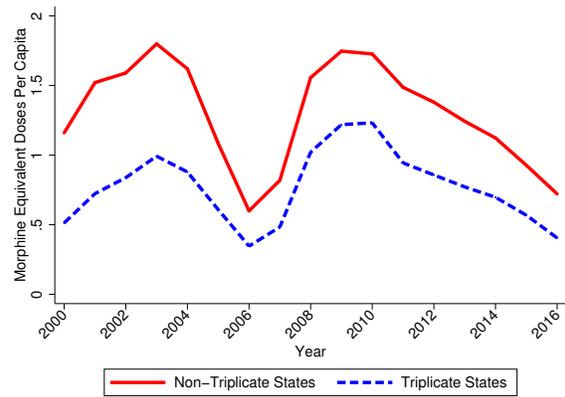
*HOUSTON LABORATORY
MARKET RESEARCH
MAY 27 1995
Purdue Frederick Company*

Source: Groups Plus (1995)

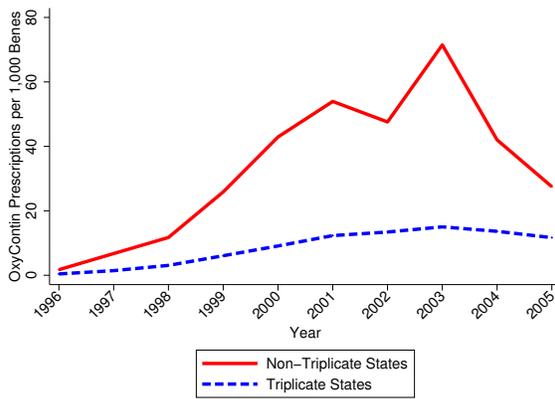
Figure A3: OxyContin Distribution and Prescriptions by Triplicate State Status



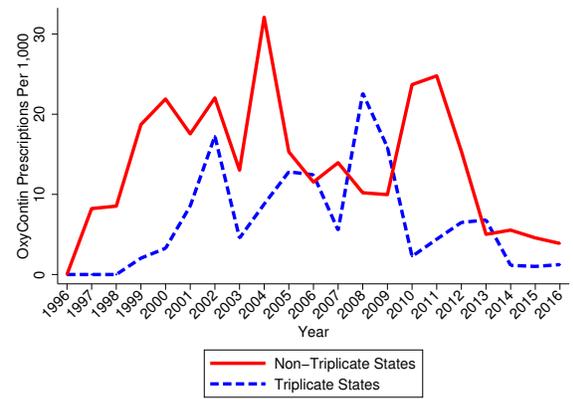
A: OxyContin Promotional Payments



B: OxyContin Distribution (ARCOS)



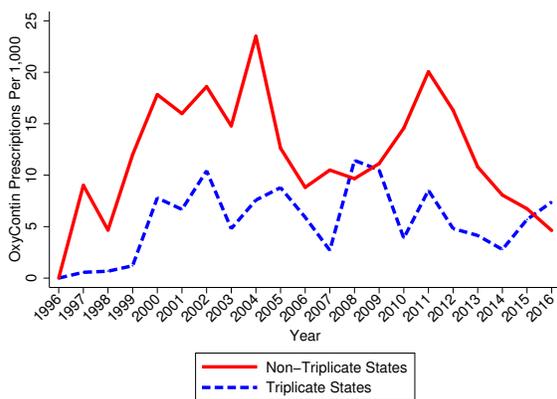
C: OxyContin Prescriptions (Medicaid)



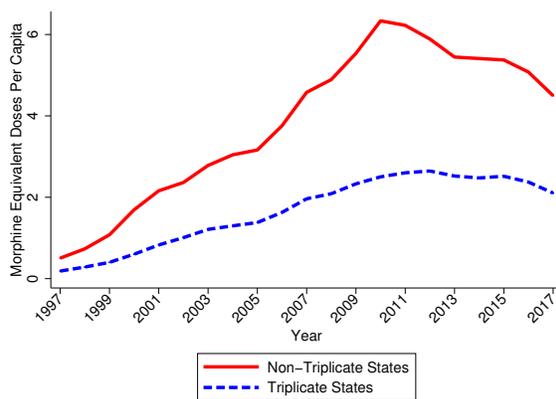
D: OxyContin Prescriptions (MEPS), Ages 25-54

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 for ages 25-54, aggregated using the MEPS survey weights. The MEPS data are the only data which allow me to stratify by age.

Figure A4: OxyContin Prescriptions and Oxycodone Distribution by Triplicate State Status (All Ages)



A: OxyContin Prescriptions (MEPS)

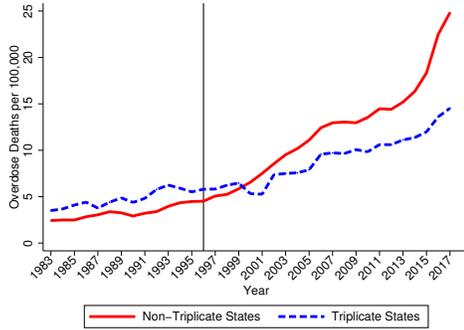


B: Oxycodone Distribution (ARCOS)

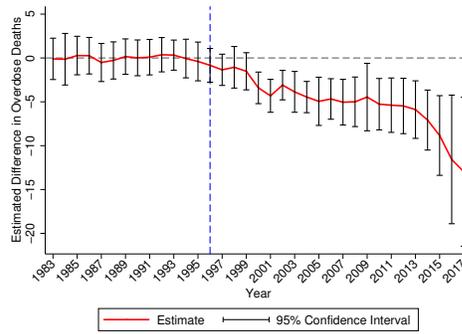
Notes: I report the number of prescriptions per 1,000 people in the MEPS in Panel A. I aggregate using the MEPS survey weights. This figure is the same as Panel D in Figure A3 but not selected on age. For Panel B, I use ARCOS data to construct oxycodone morphine equivalent doses per capita. I define a morphine equivalent dose as 60 morphine milligram equivalents. Oxycodone data are only available starting in 1997.

Figure A5: Overdose Death Trends and Event Study Estimates

Full Population

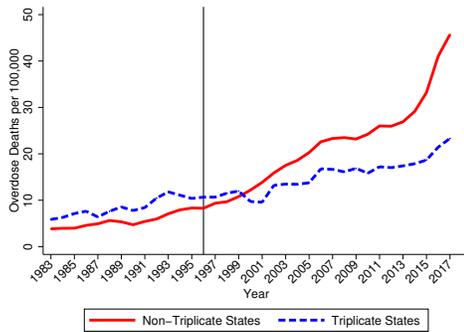


A: Non-Triplicate and Triplicate Trends

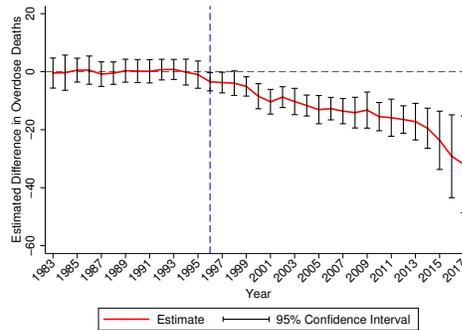


B: Adjusted Event Study

Ages 25-54



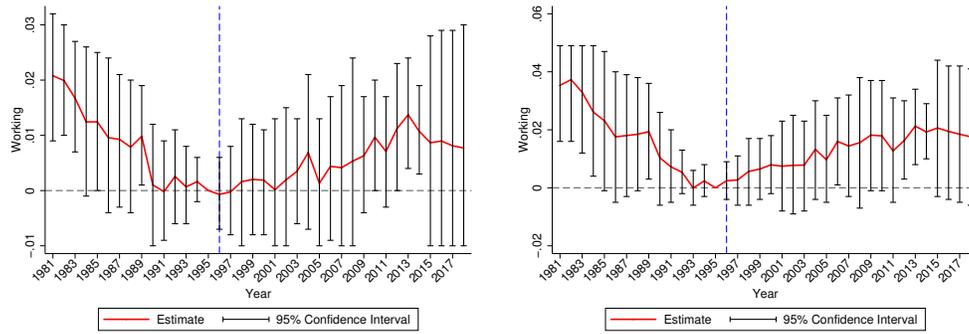
C: Non-Triplicate and Triplicate Trends



D: Adjusted Event Study

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. Panels C and D are identical to A and B, respectively, but for ages 25-54. I residualize the outcome by regressing it on state indicators, calendar month indicators, and covariates interacted with treatment status using non-triplicate states and pre-1996 data (i.e., “untreated” observations), weighted by population. I also allow the relationships between the covariates and outcome to vary over time, using lasso (with post-estimation OLS) to select the time-varying components. Covariates are same as those reported in notes for Figure 3. I report the weighted annual 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. (2022b), 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



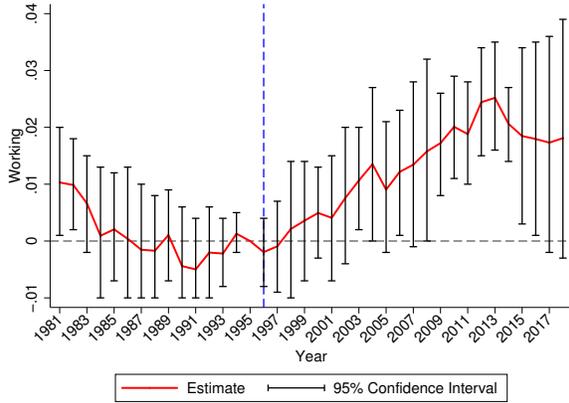
A: Labor Force Participation

B: Working

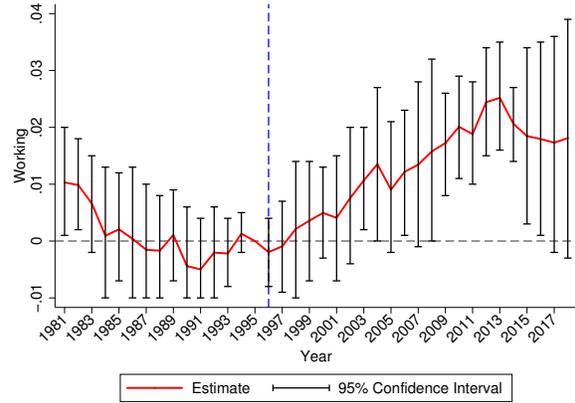
Notes: The outcome is the labor force participation rate or the share working (by month and sex). I regress the outcome on state dummies, time dummies, and Triplicate*year indicators (excluding 1995). Regressions are weighted by CPS sample weights. 95% confidence intervals are also presented and estimated using Ferman and Pinto (2019).

Figure A7: Traditional Event Study Results with Covariates

Labor Force Participation Rate

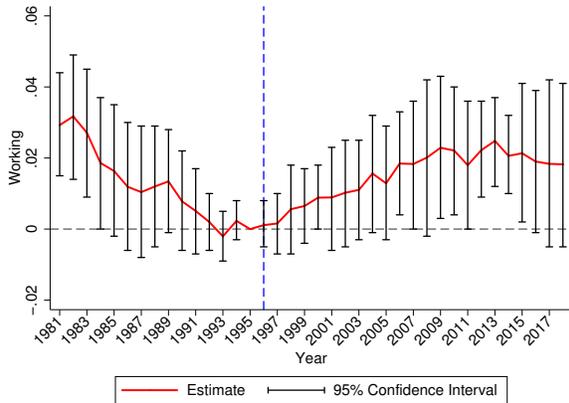


A: Baseline Controls

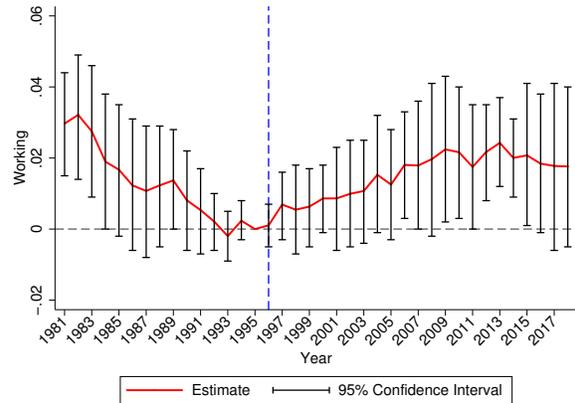


B: Lasso

Share Working



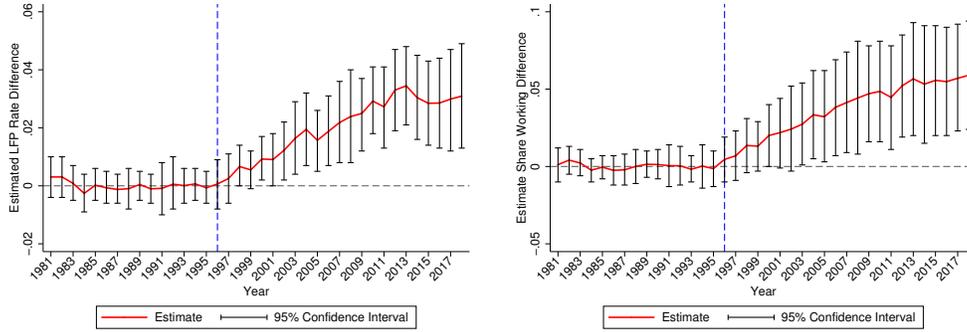
C: Baseline Controls



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, time-varying covariates, and $\text{TriPLICATE} \times \text{year}$ interactions. The coefficients on the latter set of variables are plotted. 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 (selecting from covariates interacted with year dummies as well as level and slope shifts beginning in every year 1996+). 95% confidence intervals are presented using Ferman and Pinto (2019).

Figure A8: Event Study Estimates Using State-Time as Level of Observation

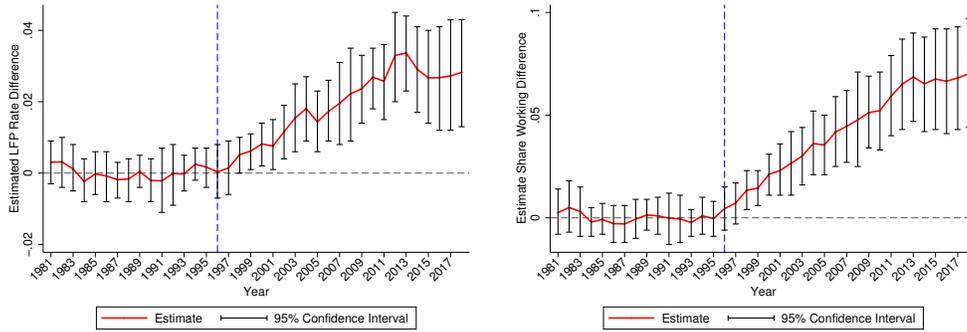


A: Labor Force Participation

B: Working

Notes: The outcome is the labor force participation rate (Panel A) or the share working (Panel B) by month and state (but not sex). I regress the outcome on state dummies, time 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 aggregated CPS sample weights) 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 CPS sample weights. The parameters associated with the covariates are permitted to vary by Triplicate status. The parameters can also vary by year – lasso is used to select these covariates. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure A9: Lasso Event Study Estimates (Penalizing All Time-Varying Covariates)

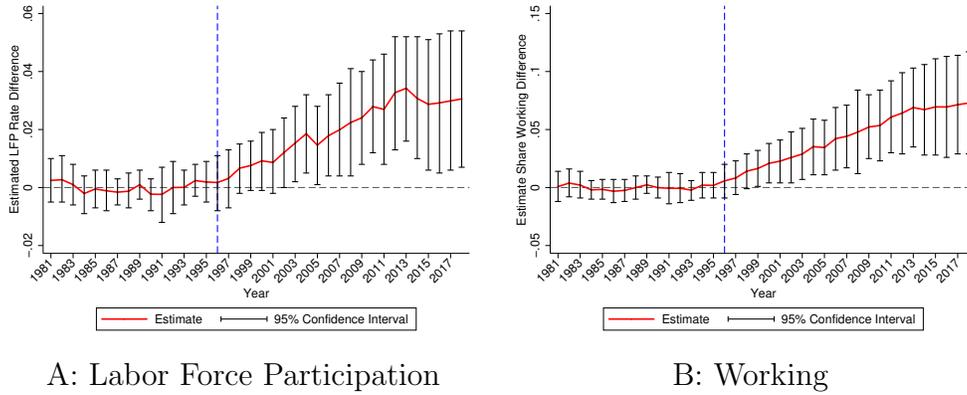


A: Labor Force Participation

B: Working

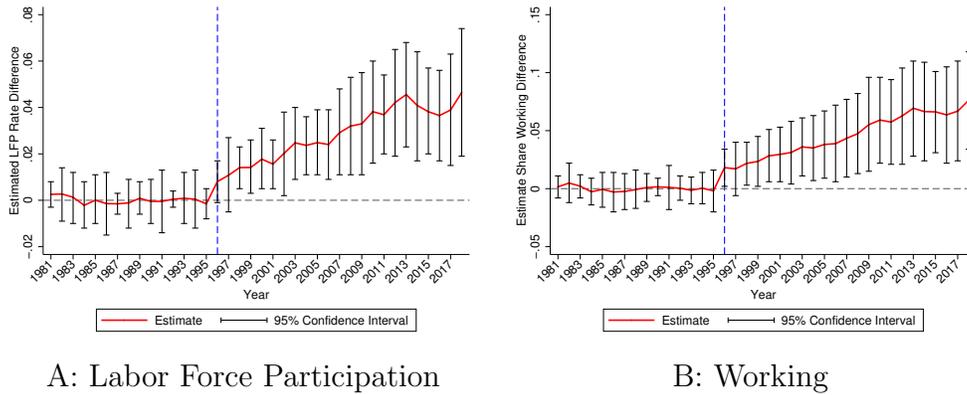
Notes: The outcome is the labor force participation rate (Panel A) or the share working (Panel B) by month, sex, and state. I regress the outcome on state*sex dummies, time*sex dummies, and time-varying covariates using pre-period data for all states and post-treatment data for non-triplicate states. Regressions are weighted by CPS sample weights. I residualize the outcome and then present annual (weighted by aggregated CPS sample weights) 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. The parameters associated with the covariates are permitted to vary by sex, Triplicate status, and year. All covariates are selected by lasso. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure A10: Lasso Event Study Estimates with Treatment-Specific Linear Trend



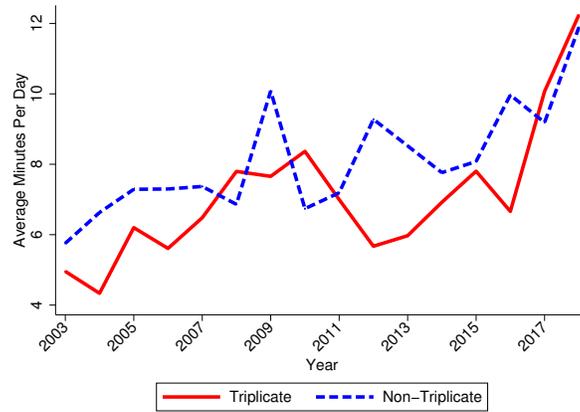
Notes: The outcome is the labor force participation rate (Panel A) or the share working (Panel B) by month, sex, and state. I regress the outcome on state*sex dummies, time*sex dummies, a linear Triplicate-specific trend term, and time-varying covariates using pre-period data for all states and post-treatment data for non-triplicate states. Regressions are weighted by CPS sample weights. I residualize the outcome and then present annual (weighted by aggregated CPS sample weights) 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. The parameters associated with the covariates are permitted to vary by sex, Triplicate status, and year. All covariates are selected by lasso. The Triplicate-specific linear trend is not penalized. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure A11: Disaggregated Data by Sex, Age, Race, and Education

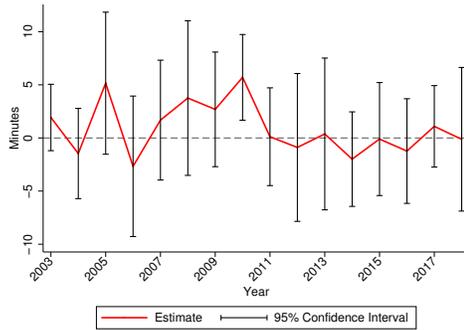


Notes: The outcome is the labor force participation rate (Panel A) or the share working (Panel B) by month, sex, age group, education, race, and state. Defining the interaction of sex, education, age group, and race as a “cell,” I regress the outcome on state*cell dummies, time*cell dummies, and time-varying covariates using pre-period data for all states and post-treatment data for non-triplicate states. Regressions are weighted by CPS sample weights. I residualize the outcome and then present annual (weighted by aggregated CPS sample weights) 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. The parameters associated with the covariates are permitted to vary by sex, Triplicate status, and year. The time component of this relationship is selected by lasso. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

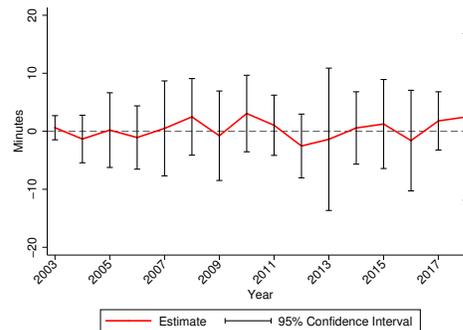
Figure A12: Number of Minutes Playing Games



A: Time Series



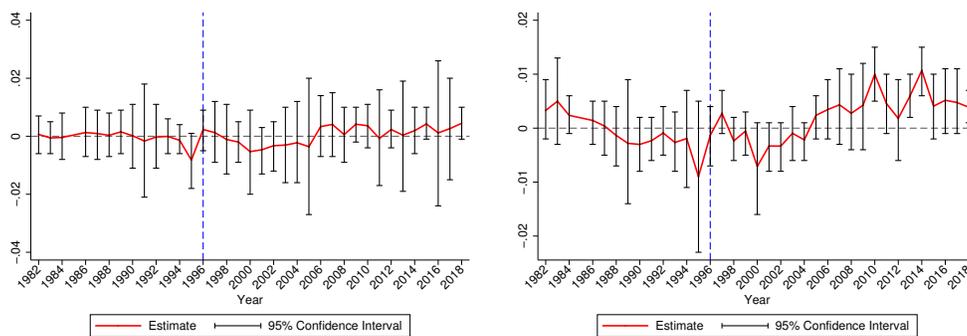
B: Adjusted Differences



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 state, 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 allow relationship between the covariates and outcome to vary over time, using lasso (with post-estimation OLS) to select the time-varying components. State indicators are not included since I do not have pre-1996 data for these outcomes. I report the weighted annual 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 A13: Migration Into and Out of State

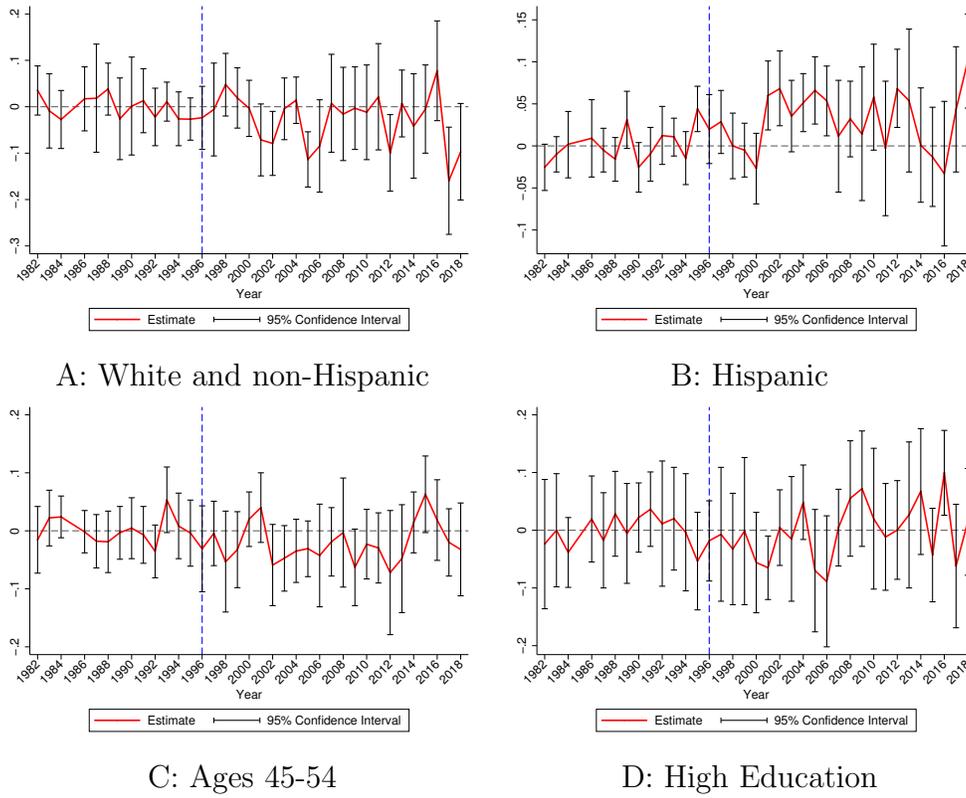


A: Out-Migration

B: In-Migration

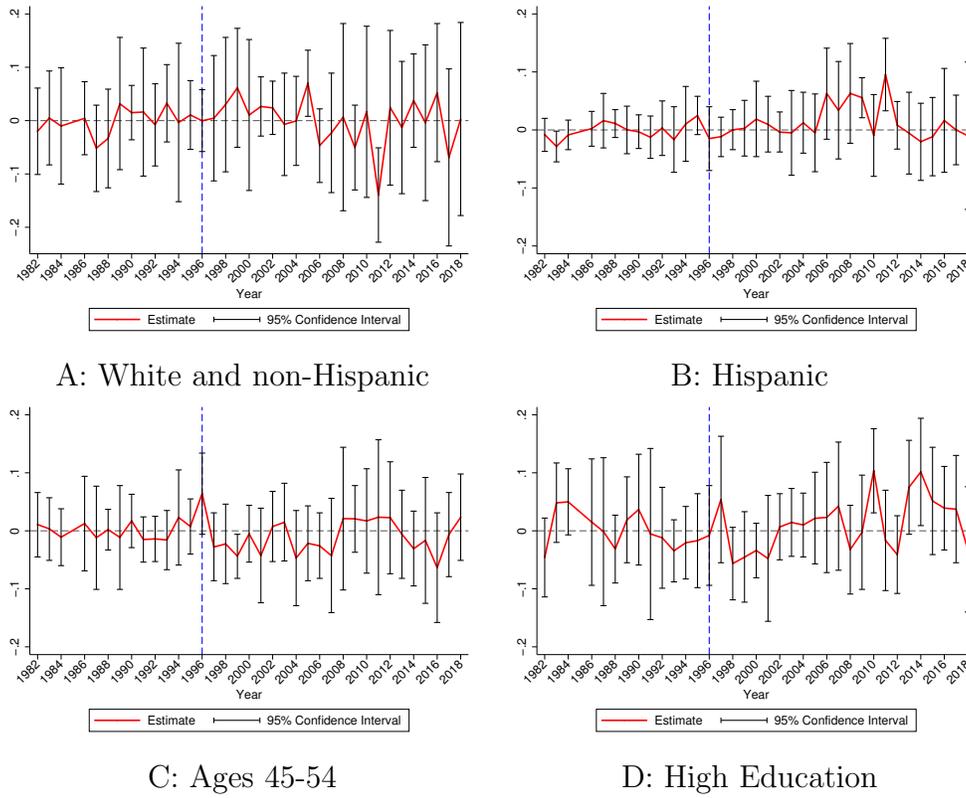
Notes: The outcome is the share of individuals migrating (or immigrating) into or out of the state (the denominator is the total population in the month-sex-state cell) by sex and month. These shares cannot be constructed for 1981 or 1985. I residualize each outcome by regressing it on state-sex and time-sex indicators using non-triplicate states and pre-1996 data (i.e., “untreated” observations), weighted by population. No additional covariates are included. I report the weighted annual means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A14: Characteristics of Individuals Migrating Into State



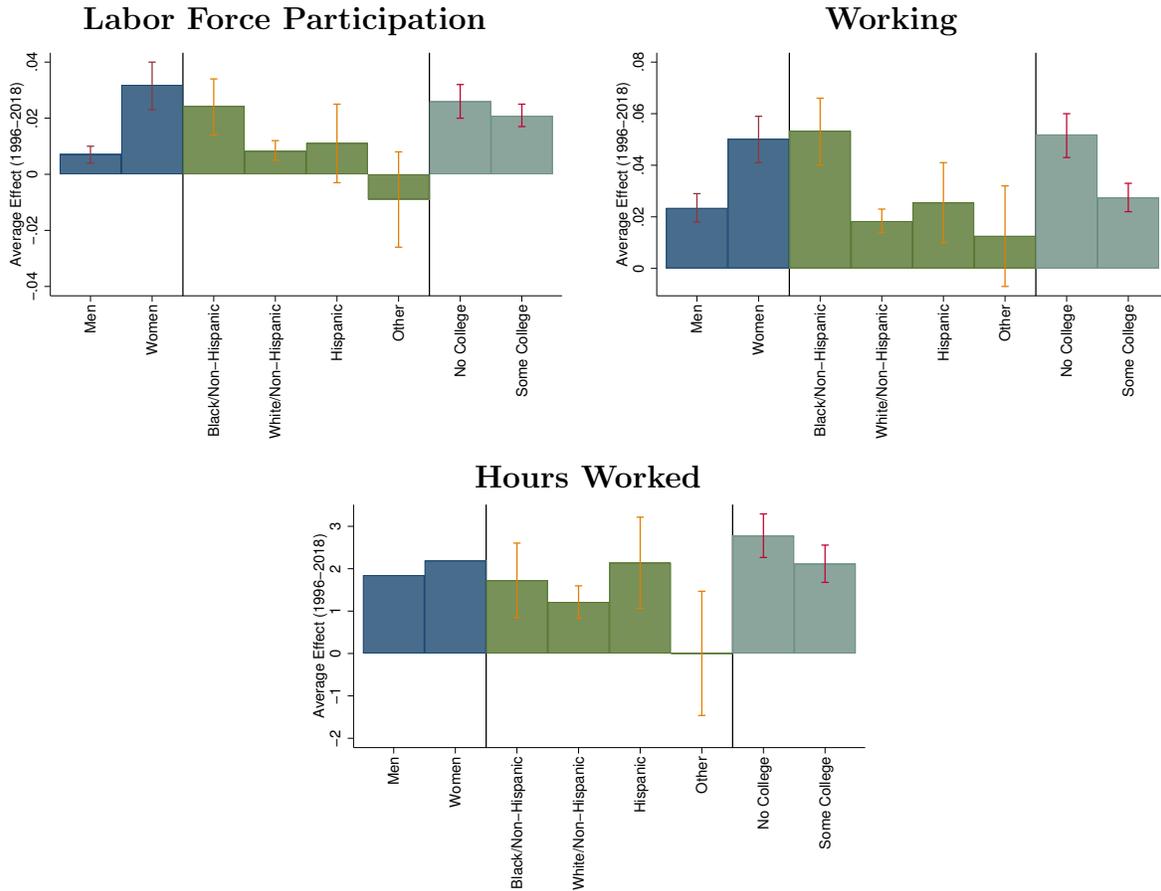
Notes: The outcome is the share of individuals migrating into the state with the listed characteristic by state-month-sex. These shares cannot be constructed for 1981 or 1985. I residualize each outcome by regressing it on state-sex and time-sex indicators using non-triplicate states and pre-1996 data (i.e., “untreated” observations), weighted by population. No additional covariates are included. I report the weighted annual means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A15: Characteristics of Individuals Migrating Out of State



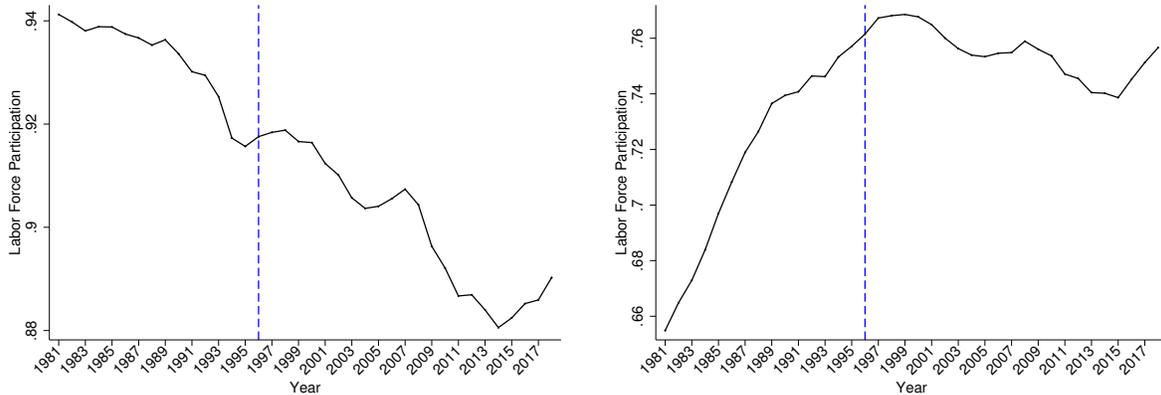
Notes: The outcome is the share of individuals migrating out of the state with the listed characteristic by state, month, and sex. These shares cannot be constructed for 1981 or 1985. I residualize each outcome by regressing it on state-sex and time-sex indicators using non-triplicate states and pre-1996 data (i.e., “untreated” observations), weighted by population. No additional covariates are included. I report the weighted annual means of the residuals for the triplicate states. 95% confidence intervals are constructed using the inference procedure discussed in the paper.

Figure A16: Demographic Heterogeneity in Labor Force Participation, Working, and Hours



Notes: Outcomes defined by listed demographic by state, month, and sex. I report the weighted average effect for 1996-2018 with 95% confidence intervals. 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 parameters are selected by lasso. 95% confidence intervals are presented and estimated using the procedure discussed in the paper.

Figure A17: Labor Force Participation Trends by Sex

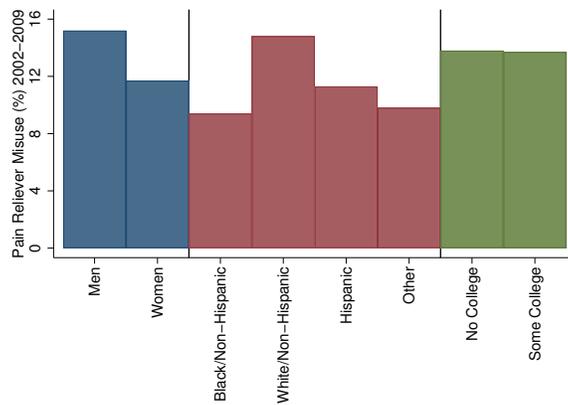


A: Men

B: Women

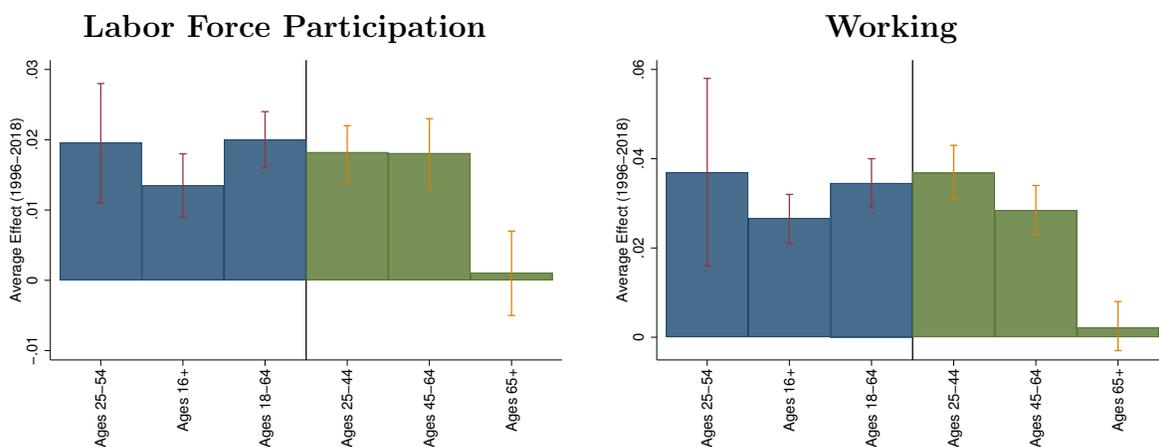
Notes: Rates of labor force participation in the CPS by sex.

Figure A18: NSDUH Pain Reliever Misuse Rates, 2002-2009



Notes: Self-reported lifetime pain reliever misuse rates by demographic. Author's calculations using NSDUH.

Figure A19: Age Heterogeneity: Labor Force Participation and Working



Notes: Outcomes defined by listed demographic by state, month, and sex. I report the weighted average effect for 1996-2018 with 95% confidence intervals. 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 White and non-Hispanic, share Hispanic, share with at least some college, and share ages 45-54. For 16+, I replace the share ages 45-54 variables with share ages 45-64 and share ages 65+. These changes have little impact on the results. 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 components 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 [-0.004, 0.002]	0.008*** [0.006, 0.011]	-0.001 [-0.004, 0.002]
2001-2010	0.007*** [0.003, 0.011]	0.011*** [0.008, 0.015]	0.007*** [0.003, 0.011]
2011-2018	0.016*** [0.011, 0.021]	0.020*** [0.017, 0.024]	0.016*** [0.011, 0.020]
1996-2018	-0.006*** [-0.010, -0.002]	0.005*** [0.002, 0.009]	-0.006*** [-0.010, -0.002]
Covariates	Baseline	Vary by Time	Selected by Lasso
B: Working			
Triplicate ×	(4)	(5)	(6)
1996-2000	0.001 [-0.003, 0.005]	0.011*** [0.007, 0.015]	0.001 [-0.003, 0.005]
2001-2010	0.014*** [0.008, 0.019]	0.015*** [0.010, 0.020]	0.014*** [0.008, 0.019]
2011-2018	0.025*** [0.019, 0.031]	0.028*** [0.023, 0.032]	0.025*** [0.019, 0.031]
1996-2018	-0.006** [-0.010, -0.001]	0.006*** [0.002, 0.011]	-0.006** [-0.010, -0.001]
Covariates	Baseline	Vary by Time	Selected by Lasso

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated using Ferman and Pinto (2019). 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, 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

A: Labor Force Participation				
	(1)	(2)	(3)	(4)
	Main	+ Bartik	+ Bartik Manufacturing	+ NTR
Triplicate ×	Results	Variable	Variable	Rate Variables
1996-2000	0.011** [0.002, 0.020]	0.011*** [0.008, 0.015]	0.012*** [0.009, 0.016]	0.013*** [0.009, 0.017]
2001-2010	0.021*** [0.006, 0.035]	0.021*** [0.017, 0.025]	0.024*** [0.019, 0.028]	0.026*** [0.022, 0.030]
2011-2018	0.034*** [0.017, 0.052]	0.034*** [0.029, 0.039]	0.038*** [0.033, 0.043]	0.042*** [0.037, 0.046]
1996-2018	0.024*** [0.012, 0.035]	0.024*** [0.019, 0.029]	0.027*** [0.022, 0.031]	0.029*** [0.025, 0.034]
B: Working				
	(5)	(6)	(7)	(8)
1996-2000	0.018** [0.001, 0.034]	0.018*** [0.012, 0.023]	0.019*** [0.015, 0.024]	0.020*** [0.016, 0.025]
2001-2010	0.028* [-0.001, 0.058]	0.029*** [0.023, 0.035]	0.033*** [0.028, 0.039]	0.039*** [0.034, 0.043]
2011-2018	0.051*** [0.017, 0.085]	0.051*** [0.046, 0.056]	0.057*** [0.052, 0.062]	0.063*** [0.058, 0.069]
1996-2018	0.034*** [0.012, 0.056]	0.035*** [0.029, 0.040]	0.039*** [0.034, 0.045]	0.044*** [0.038, 0.049]

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.3. 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. Results are similar when using lasso to select the time-varying parameters. The parameters are estimated using pre-period data from all states and post-period data for non-triplicate states. Each column adds a measure of labor demand. The final column uses the Pierce and Schott (2020) measure of exposure to permanent normal trade relations (NTR) to China. Regressions and averages are weighted by CPS sample weights.

Table A3: Accounting for China Trade Shock and Industrial Robots

A: Labor Force Participation					
China Shock			Robots		
Triplicate ×	(1)	(2)	Triplicate ×	(3)	(4)
1999	0.012*** [0.004, 0.021]	0.014*** [0.007, 0.022]	2007	0.015*** [0.006, 0.024]	0.017*** [0.007, 0.027]
2011	0.033*** [0.023, 0.042]	0.037*** [0.027, 0.047]	2014	0.027*** [0.017, 0.036]	0.029*** [0.014, 0.045]
1999+2011	0.023*** [0.017, 0.029]	0.026*** [0.020, 0.032]	2007+2014	0.021*** [0.015, 0.027]	0.023*** [0.016, 0.031]
B: Working					
China Shock			Robots		
1999	0.013*** [0.005, 0.022]	0.019*** [0.010, 0.027]	2007	0.013** [0.003, 0.023]	0.015*** [0.004, 0.027]
2011	0.041*** [0.023, 0.058]	0.053*** [0.035, 0.070]	2014	0.031*** [0.020, 0.042]	0.033*** [0.015, 0.051]
1999+2011	0.027*** [0.021, 0.034]	0.036*** [0.030, 0.043]	2007+2014	0.023*** [0.013, 0.032]	0.024*** [0.016, 0.033]
Extra Controls?	No	Exposure to Chinese Imports	No	Exposure to Robots	

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated in manner described in Section 3.3. 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 (listed in notes for Figure 3). The parameters on the covariates are permitted to vary by treatment-sex and by year-sex. These parameters are estimated using pre-period data from all states and post-period data for non-triplicate states. In Columns 1 and 2, I limit the sample to 1981-1995 (pre-period), 1999, and 2011 due to availability of the Exposure to Chinese Imports variable. Alaska and Hawaii are also excluded. In Columns 3 and 4, I limit the sample to 1981-1995 (pre-period), 2007, and 2014 due to availability of the Exposure to Robots variable. Alaska and Hawaii are excluded. Instead of reporting an estimate for 1996-2000, I report one for 1999 in Columns 1 and 2. Instead of reporting an estimate for 2001-2010, I report one for 2007 in Columns 3 and 4. Similarly, I report estimates for 2011 or 2014 instead of 2011-2018. The average post-period effect corresponds to 1999 and 2011 or 2007 and 2014. Exposure variables were aggregated to the state-level using population weights and are set to zero in the pre-period. In Columns 2 and 4, the residualization step uses two-stage least squares, and I use the respective instruments for the exposure variables. Results (in Columns 1 and 3) are shown without these additional controls to provide baseline estimates for the same time periods. Regressions and averages are weighted by CPS sample weights.

B More Details on Methods

B.1 Formal Conditions for Estimator

I assume the following conditions, which are similar to those in de Chaisemartin and D’Haultfoeuille (2020):

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

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 $E[\boldsymbol{\delta}|W_s, t]$,

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

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 $E[\boldsymbol{\delta}|W_s, t]$ which, as discussed in Section 3.2, does not require a homogeneous parameter for each covariate. In fact, this condition may be more likely to hold given additional flexibility. Given differential levels of treatment (see Section B.4), it may be important to permit heterogeneity in these parameters.

B.2 Simulations

To illustrate the problems with two-way fixed effects models as discussed in Section 3, 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:

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

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 $E[x_{st}|D_{st} = 1] = 0$.⁶⁷ 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 are bias estimates.

I show the results in Figure B1. 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 sharp 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.

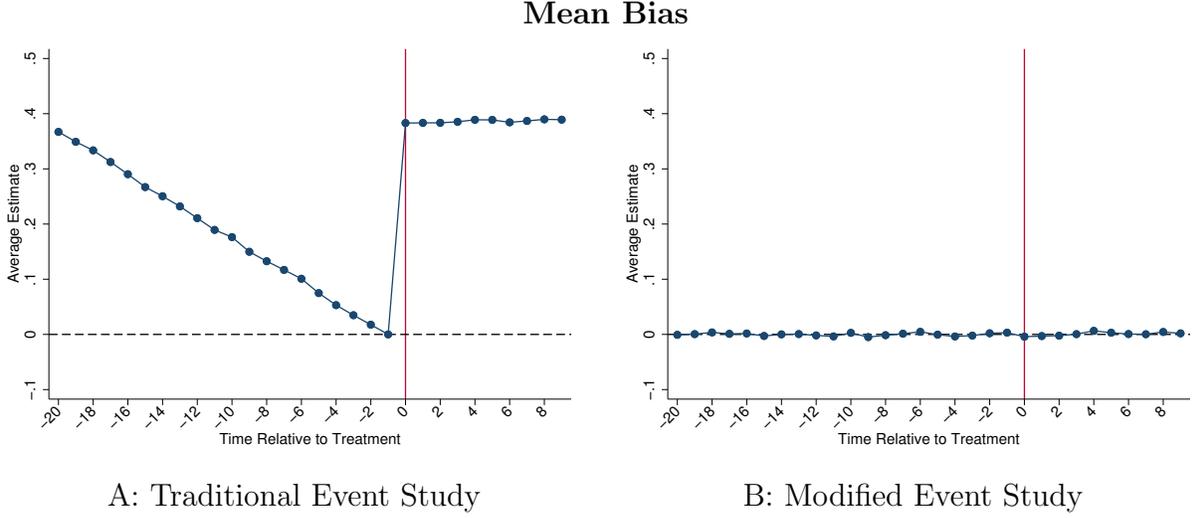
However, neither the pre-trend nor the “policy effect” beginning at time $T_0 + 1$ is real. Instead, both solely reflect trends and shocks to the *covariate* due to interactions of the treatment and the covariate.⁶⁸ This example is intentionally extreme to illustrate the underlying issue, but equation (3) shows that such bias occurs more generally.

In Panel B, I provide simulations from the proposed (in Section 3.1) “residualization approach.” The bias disappears.

⁶⁷I also let the population sizes of the units to vary for the sake of testing the inference procedure discussed later. The population sizes are constrained such that the weighted treatment effect is equal to zero.

⁶⁸In this case, not controlling for x – since it does not have an independent effect on y – produces unbiased estimates. 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.

Figure B1: Simulation Results



Notes: I plot the mean bias estimates for each time period from 10,000 simulations. See Sections B.2 and 3.1 for details.

B.3 Variance Estimation

Consider the variance of each estimate under the conditions discussed in Section 3.3, 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 $\nu_s \equiv (\nu_{s1}, \dots, \nu_{sT})$ i.i.d across s . The η_{ist} terms are i.i.d. Then,

$$\begin{aligned} \text{Var}(\hat{\beta}_b) &= \underbrace{\text{Var} \left[\sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T \frac{M_{jt} \nu_{b(j)t}}{W_{Post,Treated}} \right]}_A + \underbrace{\left\{ \sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T \left(\frac{M_{jt}}{W_{Post,Treated}} \right)^2 \left(\sum_{i=1}^{N_{b(j)t}} \left(\frac{\omega_{i,b(j),t}}{M_{b(j)t}} \right)^2 \right) \right\}}_{q_b} \sigma_\eta^2 \\ &+ \text{Var}(\mathbf{X}'_b \hat{\delta}_{w(s)t} + \sum_{j \in \mathcal{S}_1} \hat{\alpha}_{b,j} + \hat{\gamma}_t) \\ &\equiv A + Bq_b + \text{Var}(\mathbf{X}'_b \hat{\delta}_{w(s)t} + \sum_{j \in \mathcal{S}_1} \hat{\alpha}_{b,j} + \hat{\gamma}_t). \end{aligned}$$

This last relationship holds since $\text{Var} \left[\sum_{j \in \mathcal{S}_1} \sum_{t=T_0+1}^T \frac{M_{jt} \nu_{b(j)t}}{W_{Post,Treated}} \right]$ 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. The residualization step produces an estimate of the variance of $\mathbf{X}'_b \hat{\boldsymbol{\delta}}_{w(s)t} + \sum_{j \in \mathcal{S}_1} \hat{\alpha}_{b,j} + \hat{\gamma}_t$. I use the traditional cluster covariance matrix estimator (i.e., “cluster by state”) to estimate the variance of the parameter estimates. This variance estimate is valid asymptotically given a large number of untreated observations, which should hold in many applications.

Once the variance is estimated, $\widehat{\text{Var}}(\hat{\beta}_b)$ can be used to scale $\hat{\beta}_b$. The proposed bootstrap procedure, then, is valid.

B.4 Differential Levels of Treatment

In the application of this paper, all states are potentially affected by the introduction of OxyContin, but treatment levels vary. I observe useful proxies of the level of treatment by triplicate status (see Figure A3). I consider the case in which there are treated and untreated units, defined 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 pre-period data, then

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

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

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

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 $\delta_{w(s)t}$ is necessary to recover the counterfactual for the treated units and estimate the “treatment on the more treated.” Notably, however, this approach assumes that all sources of heterogeneity are observed and included in the model. Otherwise, we still estimate the causal change in outcomes between treated and untreated units, not necessarily the treatment on the treated.

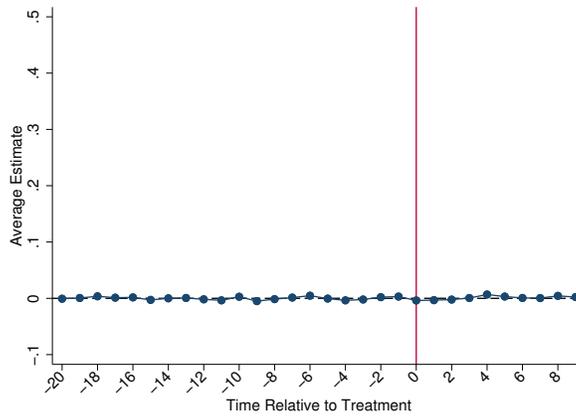
The proposed approach encourages estimating time-specific effects for the covariates in fuzzy designs. As discussed above, this approach requires including a large set of variables in the initial regression. I previously explained concerns about estimating all sources of treatment heterogeneity given a small number of treated units. 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 B.2 with a slight modification. Define

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

All units are treated in the post-period, but five are more exposed to treatment than the others. The results are provided in Figure B2. In this case, the first step of modified approach is to estimate a specification which permits 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 B2: Simulation Results: Fuzzy Design



Mean Bias

Notes: These are the results from 10,000 simulations detailed in Section B.4. 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.

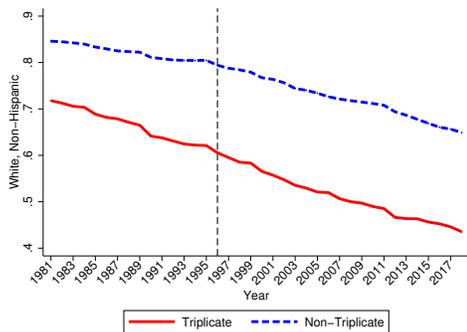
C Trends in Covariates

The unconditional labor supply trends in triplicate and non-triplicate states appear different prior to 1996 (Figure 1). 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.

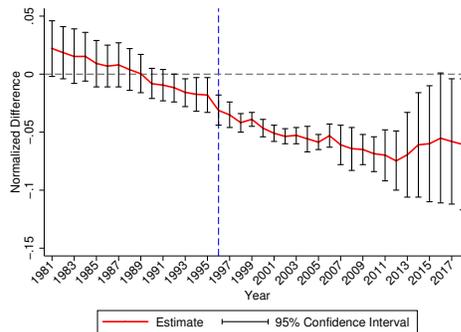
The share of the population with at least some college experience declines in triplicate states relative to non-triplicate states. This decline is relatively linear for most of the time period. Finally, I study the share of the 25-54 population which is between the ages of 45-54 since having more of 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

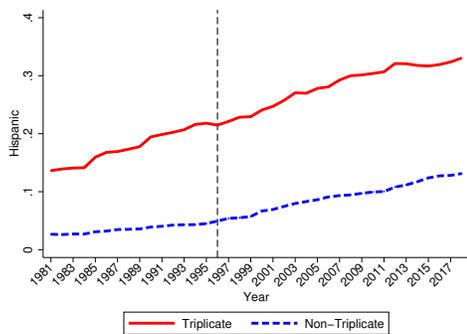


A: Time Series

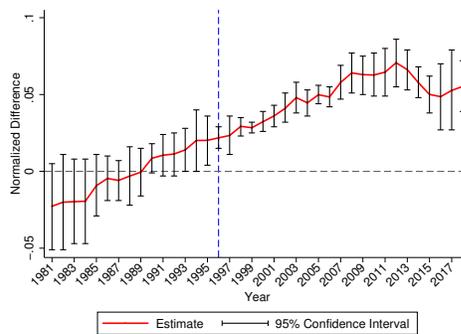


B: Differences

Hispanic

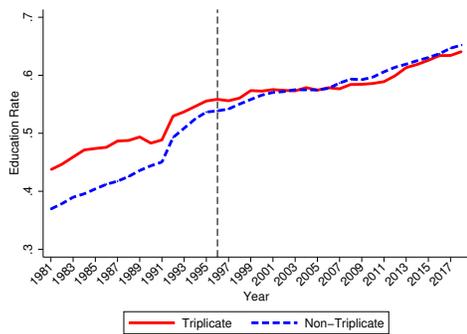


C: Time Series

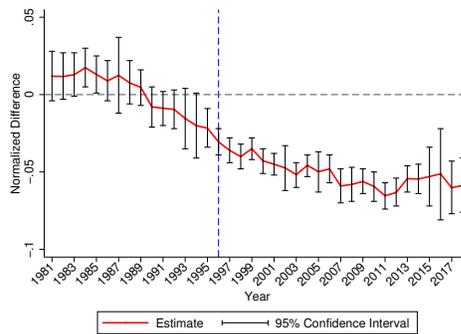


D: Differences

High Education

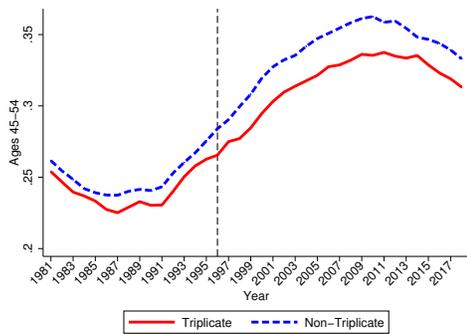


E: Time Series

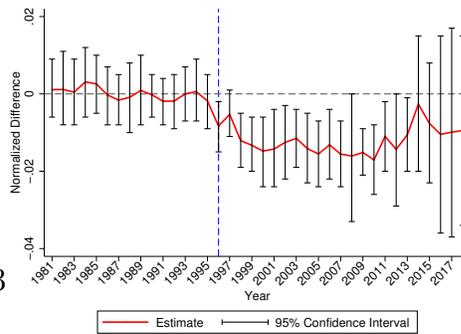


F: Differences

Ages 45-54



G: Time Series



H: Differences

Notes: I graph the covariates over time by triplicate status. I also include annual differences using the residualization approach (with state and time fixed effects) of this paper. 95% confidence intervals are generated using the inference procedure proposed in the paper.

D Alternative Inference Procedures

In this section, I use different inference procedure to estimate 95% confidence intervals for the estimates provided in Table 1. These alternate confidence intervals are provided in Table D1. The “main approach” is the approach used throughout the paper and described in Section 3.3. “Approach 2” is the same as the main approach but bootstraps *all* states, including treated states. This method ignores that the null hypothesis does not provide information about the treatment effect for any specific unit and, instead, assumes a homogenous effect. “Approach 3” is a simple permutation test which randomly selects untreated units and considers them “treated.” This method is similar to the main approach but does not make any variance adjustment. “Approach 4” is the same as “Approach 3” but randomly selects among all units – treated and untreated. “Ferman-Pinto” is the approach introduced in Ferman and Pinto (2019).

Notably, the general conclusions of the paper do not rely on the specific inference procedure used throughout the paper. Not adjusting for heteroscedasticity tends to lead to tighter confidence intervals. This is not surprising given the large population sizes of the triplicate states. Otherwise, the main approach is not consistently more or less conservative than the alternative methods. For example, the confidence intervals are tighter than the Ferman-Pinto ones for the full period labor force participation results. However, the Ferman-Pinto confidence intervals are tighter for all three time segments for the share working results.

Table D1: Different Inference Methods

A: Labor Force Participation			
Triplicate ×	(1)	(2)	(3)
1996-2000	0.005	0.011	0.005
(main approach)	[-0.002, 0.011]	[0.002, 0.020]**	[-0.002, 0.011]
(Approach 2)	[-0.002, 0.011]	[0.001, 0.022]**	[-0.002, 0.012]
(Approach 3)	[-0.004, 0.013]	[0.004, 0.018]***	[-0.003, 0.013]
(Approach 4)	[-0.001, 0.010]	[0.004, 0.018]***	[-0.001, 0.010]
(Ferman-Pinto)	[-0.002, 0.011]	[0.002, 0.020]**	[-0.002, 0.011]
2001-2010	0.017	0.021	0.018
(main approach)	[0.006, 0.028]***	[0.006, 0.035]***	[0.007, 0.028]***
(Approach 2)	[0.005, 0.030]**	[0.005, 0.036]**	[0.005, 0.030]**
(Approach 3)	[0.011, 0.023]***	[0.016, 0.025]***	[0.012, 0.023]***
(Approach 4)	[0.009, 0.026]***	[0.012, 0.029]***	[0.009, 0.026]***
(Ferman-Pinto)	[0.002, 0.032]**	[0.005, 0.036]**	[0.002, 0.033]**
2011-2018	0.03	0.034	0.031
(main approach)	[0.017, 0.044]***	[0.017, 0.052]***	[0.017, 0.044]***
(Approach 2)	[0.014, 0.047]***	[0.014, 0.055]***	[0.013, 0.048]***
(Approach 3)	[0.023, 0.037]***	[0.029, 0.040]***	[0.024, 0.037]***
(Approach 4)	[0.018, 0.043]***	[0.022, 0.047]***	[0.018, 0.044]***
(Ferman-Pinto)	[0.005, 0.055]**	[0.020, 0.049]***	[0.012, 0.049]***
1996-2018	0.019	0.024	0.020
(main approach)	[0.011, 0.028]***	[0.012, 0.035]***	[0.011, 0.028]***
(Approach 2)	[0.010, 0.029]***	[0.011, 0.036]***	[0.010, 0.029]***
(Approach 3)	[0.017, 0.022]***	[0.021, 0.026]***	[0.017, 0.022]***
(Approach 4)	[0.016, 0.023]***	[0.019, 0.028]***	[0.016, 0.023]***
(Ferman-Pinto)	[0.002, 0.036]**	[0.005, 0.042]**	[0.003, 0.037]**
Covariates	Baseline	Vary by Time	Selected by Lasso
B: Working			
Triplicate ×	(4)	(5)	(6)
1996-2000	0.011	0.018	0.012
(main approach)	[-0.003, 0.026]	[0.001, 0.034]**	[-0.004, 0.027]
(Approach 2)	[-0.006, 0.029]	[0.001, 0.034]**	[-0.006, 0.029]
(Approach 3)	[0.000, 0.023]*	[0.010, 0.026]***	[0.000, 0.023]*
(Approach 4)	[0.001, 0.022]**	[0.008, 0.027]***	[0.001, 0.022]**
(Ferman-Pinto)	[0.001, 0.022]**	[0.002, 0.034]**	[0.002, 0.022]**
2001-2010	0.034	0.028	0.034
(main approach)	[0.006, 0.062]**	[-0.001, 0.058]**	[0.005, 0.063]**
(Approach 2)	[0.008, 0.060]***	[0.002, 0.055]**	[0.008, 0.060]***
(Approach 3)	[0.025, 0.043]***	[0.020, 0.037]***	[0.026, 0.043]***
(Approach 4)	[0.019, 0.049]***	[0.015, 0.042]***	[0.019, 0.049]***
(Ferman-Pinto)	[0.009, 0.059]***	[0.007, 0.049]***	[0.009, 0.060]***
2011-2018	0.054	0.051	0.055
(main approach)	[0.027, 0.082]***	[0.017, 0.085]***	[0.026, 0.083]***
(Approach 2)	[0.017, 0.092]***	[0.013, 0.089]***	[0.017, 0.092]***
(Approach 3)	[0.047, 0.062]***	[0.044, 0.058]***	[0.047, 0.062]***
(Approach 4)	[0.031, 0.078]***	[0.029, 0.073]***	[0.031, 0.078]***
(Ferman-Pinto)	[0.031, 0.078]***	[0.030, 0.072]***	[0.032, 0.077]***
1996-2018	0.037	0.034	0.037
(main approach)	[0.015, 0.058]***	[0.012, 0.056]***	[0.016, 0.058]***
(Approach 2)	[0.015, 0.058]***	[0.011, 0.057]***	[0.015, 0.058]***
(Approach 3)	[0.032, 0.041]***	[0.030, 0.038]***	[0.032, 0.042]***
(Approach 4)	[0.030, 0.044]***	[0.027, 0.041]***	[0.030, 0.044]***
(Ferman-Pinto)	[0.006, 0.068]**	[0.006, 0.062]**	[0.006, 0.068]**
Covariates	Baseline	Vary by Time	Selected by Lasso

Notes: ***1% significance, **5% significance, *10% significance. 95% confidence intervals are estimated using the various methods discussed in this section. Estimates repeated from Table 1.

E Replication and Extrapolation Exercises

E.1 Replication

In this section, I consider the role of major labor demand shocks analyzed in the literature. I focus on the China trade shock and growth in use of industrial robots given that Abraham and Kearney (2020) find they explain a meaningful share of changes in employment rates over time. Other factors are found to have smaller (or ambiguous) impacts.

While there are several papers on the China shock, I replicate the employment analysis in Autor et al. (2019) given that it covers 2000-2014, which is most similar to the time period of interest (employment is not the main outcome of that paper, but the empirical analysis is the same). This literature constructs a measure of exposure to the China trade shock based on commuting zone industry composition and Chinese import penetration. This variable is instrumented by a related measure using non-US industry-specific growth in Chinese exports. The main specification in Autor et al. (2019) uses first differences for 1990-2000 and 2000-2014. There are 722 commuting zones, and I estimate the same specification. Autor et al. (2019) provide results for employment rates for the 18-39 age group in Table A3 (Column 1). I include their main result in Column 1 of Table E1.

To focus on the time period of interest, I next select the sample on the 2000-2014 difference and present the estimate in Column 2. In Column 3, I replicate Column 2 but construct my own measure of employment rates for the same population. The estimate is similar, suggesting that my measure is similar enough to the one used in Autor et al. (2019). In the next two columns, I focus on the 25-54 population. The estimates decrease. The estimated magnitude shrinks further to -0.060 (statistically insignificant from zero) for labor force participation.

In Panel B, I study Acemoglu and Restrepo (2020). I focus on the result in Table 7E.3 of that paper, given that it also covers 2000-2014.⁶⁹ The specification studies the differential effects of commuting zone exposure to industrial robots. This measure is constructed using commuting zone industry composition and industry-specific robot exposure. A similar measure is constructed using data on industry growth in robot use in Europe and used as an instrument.

⁶⁹The robot exposure measure is only available for certain years so Acemoglu and Restrepo (2020) extrapolate when studying different time periods.

In Column 1, I replicate the finding in Acemoglu and Restrepo (2020). Column 2 repeats this estimate since there is no time period selection for this analysis. In Column 3, I use CPS data for ages 16+ to construct an employment rate. The estimate is similar. In the final two columns, I select on ages 25-54. Again, the effect on labor force participation for this group is substantially smaller, about one-third the size of the overall employment effect.

This section suggests that the labor demand shocks did not impact the 25-54 age group as much as other age groups and had even less of an effect on labor force participation (relative to employment). This finding does not contradict the literature, and it may not be that surprising that these margins are less responsive than others. Moreover, there are likely further economic or econometric explanations for why these results are so much smaller, but a full diagnosis of the analyses in the literature is not pursued here.

Table E1: Replications

	(1)	(2)	A. Autor et al. (2019)		(4)	(5)
	Estimate in Paper	Select on 2000-2014	CPS Data (18-39)		LFP (25-54)	Employment (25-54)
Change in Import Penetration	-1.54*** (0.290)	-0.885*** (0.192)	-0.868*** (0.211)		-0.060 (0.125)	-0.322* (0.172)
	B. Acemoglu and Restrepo (2020)					
US exposure to robots	-0.339*** (0.067)	-0.339*** (0.067)	CPS Data (16+) -0.333*** (0.058)		LFP (25-54) -0.119* (0.064)	Employment (25-54) -0.255*** (0.079)

Notes: ***1% significance, **5% significance, *10% significance. Standard errors in parentheses are adjusted for state clustering. N=722. 2SLS regressions are weighted by first period population size. Panel A uses the specification, data, and controls from Table A3 (Column 1) of Autor et al. (2019). Panel B uses the specification, data, and controls from Table 7E.3 in Acemoglu and Restrepo (2020). In Column 1, I produce the same estimate in the paper. In Column 2, I select on the time period 2000-2014. In Column 3, I replace the outcome with the same measure using CPS data. In Column 4, I study labor force participation for ages 25-54. In Column 5, I study employment for ages 25-54.

E.2 Extrapolation

I follow the standard approach for converting the estimates related to “within” variation to the implications on national changes. I rely on the Column 4 estimates in Table E1. The mean value of the growth in import exposure for 2000-2014 was 1.150. The mean value of growth in exposure to industrial robots was 1.638. I multiply the estimates by these means and report them in Table 3, Panel B.

E.3 Opioid Access and Labor Demand Interactions

When trying to explain the large national reductions in labor force participation, it may be important to consider interactions to avoid double-counting. In this paper’s context,

it is possible that some of the opioid crisis impact is already quantified in effects of labor demand shocks discussed in the literature. This double-counting may occur if the opioid crisis only reduces labor supply in places also suffering negative labor demand shocks.

I consider this possibility by replicating the analysis in Section E.1. Using data and code from the original papers, I interact the main variables (and instruments) with triplicate status.⁷⁰ If interactions were important, then we would expect that triplicate states would be less impacted by labor demand shocks such that the corresponding coefficients would be positive. The results are presented in Table E2. In Columns 1 and 2, I study the China shock. The interaction with triplicate status is negative for both labor force participation and for employment, though not statistically different from zero.

In Column 3 and 4, I study the effect of exposure to robots. Again, the estimates on the interaction term are both negative and statistically insignificant. Notably, stratifying the effect into two, an overall effect and a triplicate state effect, increases the noise. However, overall, there is little evidence that there are important interactions at play. This finding is consistent with the small labor force participation effects for both of these labor demand shocks found in Section E.1.

Table E2: Interactions Between Labor Demand Shocks and Triplicate Status

	(1) LFP	(2) Employment	(3) LFP	(4) Employment
Change in Import Penetration	-0.068 (0.191)	-0.326 (0.241)		
Change in Import Penetration × Triplicate	-0.004 (0.222)	-0.042 (0.301)		
Exposure to Robots			-0.100 (0.066)	-0.232*** (0.086)
Exposure to Robots × Triplicate			-0.104 (0.354)	-0.275 (0.389)

Notes: ***1% significance, **5% significance, *10% significance. Standard errors in parentheses are adjusted for state clustering. N=722. Regressions are weighted by first period population size. Columns 1 and 2 use the specification, data, and controls from Table A3 (Column 1) of Autor et al. (2019); however, the “change in import penetration” variable and its instrument are both interacted by state triplicate status. Columns 3 and 4 use the specification, data, and controls from Table 7E.3 in Acemoglu and Restrepo (2020); however, the “exposure to robots” variable and its instrument are both interacted by state triplicate status.

⁷⁰Alternatively, I could stratify the sample by triplicate status. The results are similar.