UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ESSAYS ON EFFECTS OF PUBLIC POLICIES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Andrew Barber

December 2022

The Dissertation of Andrew Barber
is approved:

Professor Laura Giuliano, Chair

Professor Carlos Dobkin

Professor Rob Fairlie

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures v

List of Tables viii

Abstract x

Dedication xii

Acknowledgments xiii

1 Conditional Cash Lotteries Increase COVID-19 Vaccination Rates 1
- 1.1 Introduction 1
- 1.2 Methods 10
- 1.3 Data 12
  - 1.3.1 Data sources 12
  - 1.3.2 Synthetic Ohio 15
- 1.4 Results 17
- 1.5 Conclusions 24
- 1.6 References 26
- 1.7 Figures and tables 31

2 Too Legit to Quit? Labor Supply Responses to the Four-Day School Week 40
- 2.1 Introduction 40
- 2.2 Background 46
  - 2.2.1 Where and Why 46
  - 2.2.2 History of the Four-Day School Week 49
  - 2.2.3 How Might the Four-Day Week Affect Teachers? 51
- 2.3 Data 53
  - 2.3.1 OSDE Employment Records 53
  - 2.3.2 National Center for Education Statistics CCD 54

2.3.3 Four-Day School Week Panel 55
2.3.4 ACS 5-year Estimates 55
2.3.5 Emergency Credentials 55
2.4 Empirical Strategy 56
2.4.1 Event Study 56
2.5 Results 58
2.6 Conclusion 61
2.7 References 63
2.8 Figures and tables 68

3 The "Hazards" of Four-Day School Weeks: Tracking Teacher Cohorts in Oklahoma 89

3.1 Introduction 89
3.2 Data 92
3.3 Empirical Strategy 95
3.3.1 Record Linking 95
3.3.2 Duration Model 95
3.4 Results 98
3.5 Conclusion 103
3.6 References 107
3.7 Figures and tables 110

A Appendix tables and figures 119

List of Figures

1.1 Surveyed COVID-19 vaccination hesitancy by state 31

1.2 Share of population with any COVID-19 vaccination over time 34

1.3 Cumulative total COVID-19 cases recorded per 100,000 population over time 35

1.4 Cumulative total COVID-19 ICU patient-days per 100,000 population over time 36

1.5 Robustness checks of SCM estimates for the share of population with any COVID-19 vaccination by the end date, using diff. samples and specifications 37

2.1 States that have used four-day weeks (as of 2019) 69

2.2 Map of four-day school week adoption by year - National 70

2.3 Map of four-day school week adoption by year - Oklahoma 71

2.4 Attrition rate for full-time teachers in Oklahoma by year 72

2.5 Attrition rate for full-time teachers in Oklahoma by year and experience bin . . . 73
2.6 Attrition rate for full-time teachers in Oklahoma by year and locale . . . 74
2.7 Emergency teacher certificates issued by year . . . 75
2.8 Mean difference in residual attrition rate between treatment and control schools . . . 76
2.9 Mean difference in residual attrition rate between treatment and control schools . . . 77
2.10 Mean difference in residual emer. cert. rate between treatment and control schools . . . 78
2.11 Mean difference in residual emer. cert. rate between treatment and control schools . . . 79
2.12 Mean difference in residual attrition rate between treatment and control schools . . . 80
2.13 Mean difference in residual attrition rate between treatment and control schools . . . 81

3.1 Distribution of experience levels in 2006 and 2019 . . . 111
3.2 Distribution of experience levels upon leaving teaching . . . 112
3.3 Five-year retention rate of different cohorts . . . 113
3.4 Three-year retention rate of different cohorts . . . 114

A1 Estimated effects for transformations of each outcome into a common scale . . . 122
A2 Robustness checks of the synthetic control estimates for the cumulative total COVID-19 cases recorded per 100,000 population, using different samples and specifications . . . 123
A3 Robustness checks of the synthetic control estimates for the total COVID-19 ICU patient-days per 100,000 population, using different samples and specifications . . . 124
A4 Synthetic control placebo effects and rankings for other states . . . 125
A5 Manufacturers' daily market shares of first dose vaccinations over time 126
A6 Five-year retention of teacher cohorts at eventually treated schools . . . 127

List of Tables

1.1 Largest five unit weights for Synthetic Ohio using ridge augmented synthetic control for an outcome of the share of state population with any COVID-19 vaccination 32
1.2 Summary statistics for the United States, Ohio, and Synthetic Ohio . . 33
1.3 Estimation results for Ohio compared to the synthetic control 38
1.4 Aggregate estimated effects and characteristics for Ohio by time period 39

2.1 Event Study: Effect of four-day school week on teacher attrition w/ different controls 82
2.2 Event Study: Effect of four-day school week on teacher attrition using eventually treated schools 83
2.3 Event Study: Effect of four-day school week on teacher attrition using non-urban schools as controls 84
2.4 Event Study: Effect of four-day school week on emergency certificate usage 85
2.5 Event Study: Effect of four-day school week on emergency certificate usage 86

2.6 Event Study: Effect of four-day school week on teacher attrition with zero prior experience 87
2.7 Event Study: Effect of four-day school week on school's ability to hire experienced teachers 88

3.1 Cox Model: Effect of FDSWs on attrition of tenured teachers hired pre-policy 115
3.2 Cox Model: Effect of FDSWs on attrition of untenured teachers hired pre-policy 116
3.3 Cox Model: Effect of FDSWs on attrition of new teachers hired post-policy 117
3.4 Cox Model: Effect of FDSWs on attrition of new teachers hired post-policy 118

A1 State COVID-19 vaccination initiatives using conditional cash lottery incentives 120
A2 State unit weights for the ridge augmented synthetic control models . . 121

Abstract

Essays on Effects of Public Policies

by

Andrew Barber

This dissertation uses reduced-form techniques to causally answer questions of direct importance in the fields of public policy broadly, but more specifically in the areas of health economics, labor economics, and education. The first chapter examines the effectiveness of Ohio's "Vax-a-Million" vaccination campaign – a state-funded program that offered entry to a cash lottery for getting (or having already received) the vaccine for the COVID-19 coronavirus. We used an improvement upon the synthetic control method, which allowed us to generate a "Synthetic Ohio" which we could use as an untreated counterfactual. Using public health data, we find that the lottery was effective not only in boosting vaccination rates, but in also reducing COVID cases and ICU utilization. Finally, using an estimate of the high costs of ICU occupancy, we perform a back of the envelope cost-benefit analysis and find that the lottery had a benefit cost ratio of at least 10:1, saving the state of Ohio over 60 million dollars.

The second and third chapters explore teacher labor supply responses to the recent, rapid adoption of a radical change in the traditional school schedule – going from five days a week to four – in K-12 education observed throughout the country. While this policy has been adopted by 26 U.S. states, I focus my efforts on Oklahoma – a state that has seen more than 20 percent of its districts make the scheduling change

since 2010 – and use publicly-available employment records from the Oklahoma State Department of Education to examine the effect that this schedule change had on teacher retention and quality.

In the second chapter, I investigate the impact that this schedule change had on schools' ability to recruit quality teachers and to retain current (and new) faculty. Since the policy had a staggered rollout, I use an event study design to align schools in event time. I find that adoption of the policy is associated with higher retention of new teachers with zero prior experience, increased recruitment of teachers with prior experience, and a reduction in the need for emergency certifications for teachers, which serves as a proxy for improved teaching quality.

In the third chapter, I describe the labor market for Oklahoma public schoolteachers and analyze the career trajectory of teachers who are early in their careers and new at their school, some of whom are exposed to, or have selected into, the four-day schedule change. I model their tenure using a duration (or hazard/survival) model and find confirmatory evidence that the four-day week substantially improves retention.

To my wife, Mia, who bet on me, and my mom and sister, Karen and Paige.

Acknowledgments

I want to begin by thanking Jeremy West for responding when I was looking for someone to assist me with examining this crazy COVID lottery program that Ohio had just announced in early May of 2021. He found it as interesting as I did, and putting the terrible pandemic aside that generated the project, I will always have such fond memories of working together from sun up to sun down, day in and day out from when the project started in May until our revisions were accepted in December. I learned so much about the research and publication process, and I know that I benefited greatly from the peer effects.

I also want to thank my advisor, Laura Giuliano. I have tested her patience as I struggled with my research, desperately trying to squeeze something useful out of a data collection and cleaning effort that went on for far too long. Despite this, she always made time for me and it was always, always appreciated.

I also need to thank my other committee members, Carlos Dobkin and Rob Fairlie. Both were excellent professors in my core courses, and I feel very fortunate to have learned econometrics from both of them. A special thanks to Rob for being willing to read my dissertation while on sabbatical. I also want to use this opportunity to give Carlos credit for nudging me to take the stairs instead of the elevator in my second or third year. I haven't taken the elevator since, and my health is greatly improved.

Lastly, I wouldn't have even made it to UCSC were it not for the Economics Department at Sac State, so I want to thank David Lang and Mark Siegler for helping

a wandering, poker-playing engineer to learn the beauty of economics and for all the lunches and office conversations. Not many appreciate the role randomness plays in our lives, but I got especially lucky to find a department that was both willing and able to help me find a new career path.

Chapter 1

Conditional Cash Lotteries Increase COVID-19 Vaccination Rates

1.1 Introduction

Providing safe and effective COVID-19 vaccines to the public only nine months after declaring the pandemic is a remarkable feat of science and policymaking. Vaccine development is only the first hurdle, however, because community (herd) immunity requires a large share of the population to be vaccinated. Overcoming widespread reluctance to vaccinate remains a significant challenge, especially as "waning vaccine confidence has taken a toll on immunization programs across the globe" in recent years (de Figueiredo et al., 2020).

From a decision-making perspective, a person's choice to (not) be vaccinated boils down to whether their expected benefit—including altruistic benefit—outweighs their cost of vaccination. The United States and other governments have greatly reduced this cost by making COVID-19 vaccines free of charge, offering free transportation to vaccination sites, and providing easily accessible facts about the vaccines to smooth any information frictions. Despite these efforts, many people remain unpersuaded. In Figure 1.1, we use data from the U.S. Census Bureau's Household Pulse Survey to plot COVID-19 vaccine hesitancy rates by state.[1] Although there is considerable heterogeneity, ranging from 7.3 percent hesitancy in Washington, D.C. to 31.6 percent in Wyoming, it is clear that much of the U.S. population remains unwilling to vaccinate despite vaccination costs being diminished to the extent possible.

Motivated by this hesitancy, a number of states have attempted to nudge people towards vaccination by also boosting the expected benefits of being vaccinated. The most prominent form of these incentive schemes, which we refer to as a conditional cash lottery (CCL), provides people with an exclusive opportunity to win large monetary prizes only if they have received a COVID-19 vaccine.[2] A CCL is similar to a conditional cash transfer in that both incentives require people to make specific behavioral changes; however, the prize-based nature of a CCL is an important distinction.

[1] Vaccination hesitancy includes responses of "definitely not" and "probably not" as survey respondents' stated willingness to be vaccinated. Figure 1.1 uses data from February 17 to May 10, 2021, including all available vaccine hesitancy data provided by the survey prior to the intervention in Ohio that we study.

[2] We provide information about each of the state COVID-19 lottery initiatives in Appendix Table A1. In total, states have committed more than $200 million in CCL prizes for vaccinated individuals to-date.

Drawing on insights from behavioral economics, CCLs capitalize on "probability neglect," a cognitive bias wherein low-probability events are either neglected entirely or hugely overrated (Sunstein, 2002). Appealing to this behavioral bias can be particularly useful for public health objectives like vaccinations because CCL incentives should predominantly encourage people who both under-estimate communicable disease risks and over-estimate their likelihood of winning a lottery prize.

In this paper, we study the first CCL targeting COVID-19 vaccinations, which Ohio Governor Mike DeWine announced on May 12, 2021. Run by the state's Department of Health, the Vax-A-Million campaign consisted of a weekly drawing each Wednesday from May 26 through June 23, with each of the five drawings awarding one adult (18+) a prize of one million dollars and one youth (12-17) a full scholarship to any public college or university in Ohio. The total program cost was 5.6 million dollars (DeWine, 2021). A free registration provided entry into all remaining prize drawings, with the entry deadline for the final drawing ending at midnight on June 20. Importantly, only state residents who had received at least one dose of a COVID-19 vaccine prior to a drawing were eligible to win.

We evaluate how Ohio's CCL treatment affects COVID-19 vaccinations and infections by comparing how these outcomes change over time in Ohio relative to a Synthetic Ohio constructed from a weighted average of other states. To obtain this counterfactual, we employ the ridge augmented synthetic control method (SCM) developed by Ben-Michael, Feller, and Rothstein, 2021, which improves on the pioneering SCM work of Abadie and Gardeazabal, 2003 and Abadie, Diamond, and Hainmueller,

2010. Whereas the classic SCM forces all unit weights to be non-negative, potentially yielding a poor pre-treatment fit of the model, the ridge augmented version allows for negative weights by modifying the synthetic control estimation via a ridge regression outcome model. The ridge regularization parameter penalizes the distance from classic SCM weights, so this approach cleverly de-biases the synthetic control estimates while also minimizing extrapolation from untreated states' convex hull. It additionally allows for incorporating pre-treatment covariates to further improve the model fit.[3]

Our study uses daily state-level data primarily from the U.S. Department of Health and Human Services and the Centers for Disease Control and Prevention (CDC). Our outcomes are COVID-19 vaccinations, COVID-19 cases (positive tests), and COVID-19-related intensive care unit (ICU) patient-days. Because state populations vary and our outcomes of interest grow monotonically over time (e.g., the total count of vaccinated people), we specify each dependent variable cumulatively as a ratio to state population (e.g., the vaccinated share of state population). We also incorporate several covariates that capture residential, political, behavioral, and supply-side factors related to accessibility of or preferences about the vaccines. Our study period spans from February 19, 2021, the earliest comprehensive data on vaccinations, to July 18, 2021, 28 days after Ohio's lottery entry ended.

We find an increase in COVID-19 vaccinations in Ohio that begins almost immediately after the Vax-A-Million announcement and persists past the final prize

[3] As we discuss and show below, our empirical findings are robust to instead using the classic synthetic control model, to including no covariates, and to a large variety of alternative model specifications.

drawing. Relative to the synthetic control, the program causes a 0.7 percentage points (1.5 percent) increase in the share of state population receiving at least a first dose of a COVID-19 vaccine by the program's end date, with most of this effect occurring within two weeks of the announcement. In levels, this amounts to about 82,000 people who were persuaded to vaccinate by the CCL incentive, implying an average program cost of 68 dollars per "complier." For context, this cost-per-complier is less than the 80 dollars in direct costs that the federal government pays a healthcare provider to fully vaccinate one person (U.S. CMS, 2021).

In turn, we find that this heightened level of vaccination subsequently reduces the spread and impact of COVID-19 within the state. Using the same framework, we estimate that Ohio's program reduces case volumes by around 125 per 100,000 population (1.3 percent) and COVID-19-related ICU patient-days by around 41 per 100,000 population (2.5 percent) by the end of our study period. In aggregate, these estimates correspond to nearly 15,000 cases and 5,000 ICU patient-days prevented (approximately 325 patients). Moreover, because of the exponential nature of disease transmission, these estimates are likely to greatly understate the total longer-run reductions relative to counterfactual.[4]

Inference is almost always the most challenging aspect of using a synthetic control method, given that there is only a single treated unit. Following Abadie, Diamond, and Hainmueller, 2010, many SCM studies resort to a form of cross-sectional

[4]We do not attempt to model the long-run effective reproduction number of infections prevented. Such an exercise is complicated because of new genetic variants of SARS-CoV-2 and because the basic reproduction number (R_0) for compliers encouraged by the incentive likely differs from that of the broader population.

permutation inference by estimating "placebo effects" for untreated units—comparing the ratio of the post-treatment root mean squared prediction error (RMSPE) and pre-treatment RMSPE of each of these estimates to that for the treated unit. We also conduct this RMSPE-based inference, finding strong support for Ohio's candidacy as an outlier among placebo-treated states. However, as Abadie, 2021 discusses, this approach suffers from some "complications," most notably that it only "reduces to classical randomization inference when the intervention is randomly assigned, a rather improbable setting." In our study context, this condition is equivalent to the assumption that Ohio is a randomly-selected state to have implemented the first CCL for COVID-19 vaccinations, which—as supported by Figure 1.1 above—is unlikely to be the case.

Fortunately, substantial progress has been made recently in the econometrics and statistics literature pertaining to inference for the SCM. These modern approaches extend the conformal prediction techniques of Vovk, Gammerman, and Shafer, 2005 to leverage the pre-treatment time series variation of the treated unit and synthetic control, rather than relying on cross-sectional comparisons. For this study, we primarily conduct inference for our estimates using the jackknife+ method developed by **Barber˙jackknifeplus** and applied to the ridge augmented SCM by Ben-Michael, Feller, and Rothstein, 2021. Jackknife+ inference operates through a leave-one-out approach of iteratively dropping each pre-treatment time period and re-estimating the model, and then uses this range of estimates to form confidence intervals for each post-treatment time period. We additional present results using the conformal inference method of Chernozhukov, Wüthrich, and Zhu, 2021, which yields similar findings. The key assumption for both approaches is that, under the null hypothesis, the distribution of differences between the treated unit and control unit is stationary over time, i.e., that time periods or residuals are exchangeable. In our study, this assumption is visually supported by examining the plots we provide of daily pre-treatment differences in outcomes between Ohio and Synthetic Ohio over time.

Our paper makes several contributions to the health economics literature. Most directly, we provide one of the only examinations of a large-scale conditional cash lottery. Although lottery-based incentives have been used conceptually for over sixty years to encourage behavior change related to public health (British Medical Journal, 1957), the limited empirical evidence is somewhat mixed and focuses primarily on smaller interventions in clinical trials or field experiments.[5] Moran et al., 1996 find that a lottery-based gift card incentive is less effective than an educational brochure at encouraging influenza vaccinations. Volpp, John, et al., 2008; Volpp, Loewenstein, et al., 2008 test small CCL incentives for losing weight and for anticoagulant drug adherence, finding success in encouraging behavioral change. Thirumurthy et al., 2016 show that offering lottery prizes does not increase voluntary medical circumcision by men in Kenya. In two recent field experiments, Goette and Stutzer, 2020 find that blood donors in Switzerland are more likely to donate again when offered a lottery ticket and Björkman Nyqvist et al., 2018 find that a CCL for safer sexual behavior in Lesotho reduces HIV incidence. Additionally, this latter study demonstrates that lottery-based incentives primarily appeal to

[5] A larger related literature examines conditional cash transfers for public health objectives including vaccinations (e.g. Barham and Maluccio, 2009). As noted above, the uncertainty in CCLs is a key distinction.

individuals with greater risk tolerance, further supporting a mechanism of probability neglect.

Three concurrent research studies also examine aspects of Ohio's Vax-A-Million program. Walkey, Law, and Bosch, 2021 conduct an interrupted time series study of Ohio versus the United States during the few weeks surrounding the lottery announcement, concluding that Ohio's program does not increase vaccination rates. However, Ohio's vaccination rates track poorly with national rates during the pre-treatment period—a factor motivating our synthetic control identification strategy. Lang, Esbenshade, and Willer, 2021 use the classic SCM to study how Ohio's program affects the share of *fully* vaccinated residents, finding no effect. This null effect could be because lottery eligibility only required a single dose rather than full vaccination. In addition, their study stops at the final lottery drawing, weeks before many lottery-eligible participants could have obtained a second dose of a vaccine series, which require 21 or 28 day gaps between doses. The study also includes all states in the donor pool, even those that also started CCLs during Ohio's treatment period, which might induce attenuation bias. Finally, M. E. Brehm, P. A. Brehm, and Saavedra, 2021 use county-level data from Ohio, Indiana, Michigan, and Pennsylvania to conduct pooled SCM and state-border difference-in-differences estimations of how Ohio's program affects the number of of first dose vaccinations during the treatment period. The study finds an effect on vaccinations that is very similar in magnitude to that we show here.

To our knowledge, we provide the only evidence about how Ohio's CCL initiative ultimately affects COVID-19 infections and hospitalizations, through a mechanism of increased vaccinations. We believe this is a critically important dimension for evaluating this novel policy instrument because it allows for a comprehensive cost-benefit analysis of what has appeared, at least initially, to be a controversial program (Buchanan, 2021). Collectively, our estimates indicate that—by nine weeks after the announcement of the program—Ohio's CCL prevents at least one COVID-19 infection for every six vaccinations that the lottery successfully encourages and prevents at least one ICU patient-day for every 17 vaccinations that it encourages. As noted above, these effects are only growing stronger over time due to the exponential nature of disease transmission. Based on Di Fusco et al.'s (2021) values for COVID-19-associated ICU expenses, our estimates imply a reduction in hospital charges of around 65 million dollars, a social benefit that is an order of magnitude larger than the 5.6 million dollar cost of the program. Thus, even without including any other short- or long-run benefits from reducing COVID-19 incidence, Ohio's CCL program passes an economic cost-benefit analysis with flying colors.

With these findings, we also contribute to the literature evaluating how COVID-19 vaccination rates affect community infections. We provide evidence specific to the subset of the population that is persuaded to vaccinate only by a lottery-based financial incentive, in contrast to evidence from vaccinations of people motivated by altruistic reasons or seeking self-protection from the virus. Against a backdrop of increasing hesitancy globally towards vaccinations, this distinction could be quite valuable for public health policymakers. Inspired by Ohio's approach, at least nineteen other state governments have followed suit with their own "vaccination lotteries," with substantial heterogeneity in programmatic design. We leave it to future research to provide further insights about what, specifically, serves as the optimal form of conditional cash lottery to encourage COVID-19 vaccinations.

1.2 Methods

Our primary empirical strategy to estimate the effects of Ohio's lottery incentive treatment is the ridge augmented synthetic control method (Ben-Michael, Feller, and Rothstein, 2021). At its core, this approach compares outcomes in Ohio to outcomes in other states over time. As we show below, Ohio's vaccination rates do not track closely with overall rates in the United States even in the weeks before the Vax-A-Million program announcement, such that a simple average across other states serves as a poor counterfactual. By using the synthetic control method (SCM) to form a weighted average of the untreated states, we obtain a much better counterfactual for Ohio. Here, we provide only a basic illustration of the method, referring interested readers to Ben-Michael, Feller, and Rothstein, 2021 and Abadie, 2021 for additional details.

For panel data on states i across time periods t, denote the outcome variable as $y_{i,t}$. We are interested in the treatment effect, τ^{CCL}, of a conditional cash lottery on this outcome. Suppose for simplicity that only Ohio is ever treated and that there is only a single post-treatment period when $t = T$. In a potential outcomes framework, we can express Ohio's post-treatment outcome as $y_{Ohio,T} = y_{CF,T} + \tau^{CCL}$, where $y_{CF,T}$

is the counterfactual at time T. The essence of the SCM is to form this counterfactual for post-treatment Ohio as a weighted-average of the outcome for untreated states:

$$y_{CF,T} = \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} y_{i,T} \qquad \text{where} \qquad \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} = 1 \qquad \text{and} \qquad \hat{\gamma}_i^{scm} \geq 0 \;\; \forall i \tag{1.1}$$

The classic SCM determines these $\hat{\gamma}_i^{scm}$ weights by minimizing the differences between Ohio and the counterfactual for the outcome in pre-treatment time periods, $t < T$, as well as optionally minimizing differences in covariates between Ohio and the counterfactual. Denote the vector of pre-treatment outcomes and covariates for a state as X_i. In an ideal setting, the SCM weights would yield a near-perfect counterfactual for Ohio, i.e., $X_{Ohio} \approx \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} X_i$. In practice, it may not be feasible to determine a set of non-negative weights such that $X_{Ohio} \approx \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} X_i$, and the synthetic control will yield a poor $y_{CF,T}$ counterfactual.

To improve the quality of the counterfactual, the ridge augmented SCM layers a ridge regularized linear model onto the classic SCM:

$$y_{CF,T} = \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} y_{i,T} + \left(X_{Ohio} - \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} X_i \right) \cdot \hat{\eta}^{ridge} \tag{1.2}$$

where $\hat{\eta}^{ridge}$ are coefficients from a ridge regression of the untreated states' post-treatment outcomes $y_{i,T}$ on centered pre-treatment outcomes X_i, with a tuning parameter that limits the degree of extrapolation from the untreated states' convex hull. If the quality of the classic SCM counterfactual is very good, then $X_{Ohio} - \sum_{i \neq Ohio} \hat{\gamma}_i^{scm} X_i$ is close to zero and the ridge augmented SCM is virtually equivalent to the classic SCM. For non-trivial cases, Ben-Michael, Feller, and Rothstein, 2021 demonstrate how Equation (1.2) can be expressed as:

$$y_{CF,T} = \sum_{i \neq Ohio} \hat{\gamma}_i^{aug} y_{i,T} \qquad \text{where} \qquad \sum_{i \neq Ohio} \hat{\gamma}_i^{aug} = 1 \tag{1.3}$$

Although $\hat{\gamma}_i^{aug}$ can take negative values, unlike $\hat{\gamma}_i^{scm}$, the method directly penalizes the distance between the ridge augmented SCM weights and the classic SCM weights using the tuning parameter. Thus, the ridge augmentation de-biases the classic synthetic control estimates to improve the quality of the counterfactual for Ohio while also minimizing extrapolation. Empirically, we find that the classic SCM improves the quality of the counterfactual for Ohio relative to a simple average of untreated states, and the ridge augmented SCM further improves the quality of the Synthetic Ohio to support causal inference.

1.3 Data

1.3.1 Data sources

Our study compiles data from a variety of public sources. Data on COVID-19 vaccinations are provided by the U.S. Centers for Disease Control and Prevention (CDC), which aggregates information from state and local health departments. Specifically, this dataset includes the number of vaccines administered daily in each state.

These vaccination counts are separated by manufacturer: Janssen (Johnson & Johnson), Moderna, and Pfizer. The daily counts are also separated into first and/or final dose vaccinations. Although some vaccination data is sparsely available for earlier time periods, February 19, 2021 is the first date on which all states report comprehensive data, and we use this date to start the panel used in our analysis. Inspecting the data, there are some clear inaccuracies in the daily counts of vaccinations—such as a negative amount of vaccines being administered—anomalies which are also discussed in the CDC's data documentation.[6] Most of these errors are simply misattribution of some vaccinations to a date the day before or after the vaccines were actually administered. To correct for these inaccuracies, we smooth vaccination counts for a small number of state-weeks containing these "outliers" using an approach that preserves the cumulative vaccination counts for each state in each week but reduces artificial noise from erroneous data classification.[7]

To assess COVID-19 outcomes, we use data from the CDC for the total cases recorded in each state by date and we use data from the U.S. Department of Health and Human Services for the total volume of hospital intensive care unit (ICU) patients with COVID-19 by state-date. We use an outcome measure of ICU patient-days rather than patient counts because the data report the daily number of ICU-hospitalized patients,

[6]Discussion of data anomalies and other data reporting considerations is provided in the CDC's data documentation available at www.cdc.gov/coronavirus/2019-ncov/vaccines/distributing/about-vaccine-data.html.

[7]Specifically, we tag outlier observations using a criterion of daily vaccination volumes being greater than twice the state's seven-day moving average, adjusted for state-specific day of the week. For each state-week containing an outlier, we reallocate the total weekly volume across the days of that week using state-specific day-of-week weights. This approach leaves total vaccination counts in each state-week unaffected, reallocating only within state-week. Less than three percent of observations are outliers necessitating these corrections.

who typically stay multiple days (an average of 14.7 days per Di Fusco et al., 2021, with substantial heterogeneity). Because state populations vary and our outcomes of interest grow monotonically over time (e.g., the total COVID-19 cases recorded), we specify each dependent variable cumulatively as a ratio to state population (e.g., the total cumulative COVID-19 cases per 100,000 population). We do so using state population data for 2020 from the U.S. Census Bureau. We do not evaluate deaths from COVID-19 because this outcome is statistically under-powered—there were a total of six deaths per 100,000 population in Ohio during the post-treatment period.

Although we also present results from models without covariates, to improve the model fit we incorporate some pre-treatment state-level covariates related to accessibility of or preferences about the vaccines. We include population density (Census Bureau) and gross domestic product per capita (Bureau of Economic Analysis) as rough proxies for variation across states in the living circumstances and economic activity that could influence vaccine hesitancy, either directly or through the heterogeneous impact of COVID-19 across states during the pandemic. We include 2020 Republican presidential vote share because political leaning has been linked to vaccination hesitancy (Ivory, Leather, and Gebeloff, 2021). States' pre-pandemic influenza vaccination rates for 2019 from the Centers for Medicare and Medicaid Services are included to capture variation in more general propensities towards vaccination. We use Google's Community Mobility Reports indices to capture variation in behavior as reflected in visits to different types of places during the pre-treatment period.[8] Finally, we compute distance measures of state population to vaccination sites using Census Block Group population centers and the locations of all COVID-19 vaccination sites in the U.S. from www.vaccinatethestates.com. We use the median distance and 95th percentile distance of population to vaccination sites to proxy for differences in vaccine accessibility.

1.3.2 Synthetic Ohio

We use the ridge augmented synthetic control method to determine state unit weights for Synthetic Ohio. The donor pool includes all states that did not initiate their own lottery schemes for COVID-19 vaccinations before the end of Ohio's Vax-A-Million program (i.e., we exclude states listed in Appendix Table A1, sans Michigan and Missouri). In some of our robustness checks, we relax this requirement to include all 50 states and Washington, D.C.

Table 1.1 shows the largest five unit weights for Synthetic Ohio, using an outcome of the share of state population with any COVID-19 vaccination. Appendix Table A2 shows the full set of unit weights for Synthetic Ohio, which includes some negative weights as discussed in the methodology section above. Other than Idaho, the five states that contribute the most towards Synthetic Ohio are all also located in the Midwest Census Region: Wisconsin, Kansas, Michigan, and North Dakota. Like Ohio,

[8] Google's Community Mobility Reports provide proxies for movement over time across six categories of places: retail and recreation, groceries and pharmacies, workplaces, residential, parks, and public transit. Each index can range from −100 to 100. Google defines these using movement of people's cell phones to different places, with the baseline zero-value for each index set during January 3 through February 6, 2020.

these five states generally have high levels of surveyed pre-treatment vaccine hesitancy, shown in Figure 1.1 discussed above (Michigan, the apparent exception, has a hesitancy rate of 14.7 percent, just below the bin cutoff).

Table 1.2 presents summary statistics for the United States, Ohio, and Synthetic Ohio. Panel [A] shows shows values for the dependent variables during the pre-treatment time period(s) indicated. Panel [B] shows values for state covariates during 2019 or 2020, as indicated, or over the full pre-treatment period of February 19 through May 11, 2021. In the first row, Ohio's vaccination rate closely matches that of the U.S. overall as of April 2, 2021. However, the U.S. vaccination rate greatly outpaces Ohio's during the subsequent pre-treatment weeks such that, by the May 12 lottery announcement, Ohio's vaccinated population share lags the state average by almost four percentage points, a pattern that is shown even more clearly in the time series graphs presented in the next section. In contrast, the weighted average vaccination rate of the states that comprise Synthetic Ohio remains much closer to the rate in Ohio. The remaining rows of Table 1.2 also show a clear improvement of the counterfactual by using a synthetic control rather than a simple average of untreated states. The three outcomes we evaluate and all state covariates are (often much) closer between Ohio and Synthetic Ohio than between Ohio and the United States' average.

1.4 Results

We begin our analysis by examining vaccinations. We focus on first dose vaccination rates rather than fully-vaccinated rates because Ohio's Vax-A-Million program only required a single dose of any COVID-19 vaccine for eligibility.[9] Although completing a vaccine series provides more protection against the virus, even a single dose of the Moderna or Pfizer vaccine has been found to provide substantial immunity (Dagan et al., 2021). Regardless, as we discuss and show below, there is no difference in the vaccination series follow-up rates in Ohio from before compared to after the Vax-A-Million program.

Figure 1.2 plots first dose vaccination rates over time. As shown in Panel (a), there are virtually no differences between Ohio, Synthetic Ohio, and the entire United States in the share of vaccinated population until early April. Then, there is a structural break between Ohio and the rest of the country, with this gap growing over time. In contrast, the vaccination rate continues to be nearly identical between Ohio and the synthetic control until Ohio's conditional cash lottery treatment begins on May 12. This figure highlights the importance of using the synthetic control method because it shows how poorly a simple average of other states would serve as a counterfactual to Ohio, despite having had similar vaccination behavior in earlier months prior to the start of the treatment.[10]

[9] A single vaccine dose is also the requirement for all other state initiatives shown in Appendix Table A1 other than Massachusetts, which requires full vaccination for eligibility.

[10] While states had slightly different vaccination eligibility timelines, every adult in the United States was eligible for a COVID-19 vaccine by no later than April 19, 2021 (Biden, 2021).

The magnitude of the difference across time between Ohio's first-dose vaccination rate and its counterfactual is shown in Panel (b) of Figure 1.2. The introduction of the lottery incentive causes an almost immediate increase in vaccination rates in Ohio compared to the synthetic control—the difference is larger only three days after treatment begins than on any date over the nearly three months pre-treatment. Following the announcement of the lottery, the estimated treatment effect increases sharply over the first two weeks before leveling off, which also coincides with the timing of the first prize drawing. Of more general interest, the alacrity with which compliers respond suggests that a long treatment window may not be required to maximize the efficacy of a CCL to change behavior.

Although increasing vaccination uptake is the most direct effect of Ohio's program, ultimately the objective is to reduce COVID-19 infections, hospitalizations, and deaths. We show the first of these downstream effects of Ohio's increased vaccination rate in Figure 1.3. Given that COVID-19 vaccines take approximately 14 days to demonstrate partial efficacy, one would not anticipate to see declining infection rates until at least the end of May (Dagan et al., 2021). Furthermore, due to the exponential nature of viral transmission, any effect that is observed should grow over time as each infection prevented then also prevents additional cases. This expected pattern matches the evidence shown in Figure 1.3. There is little effect of Ohio's program on COVID-19 cases until early June, but then the cumulative difference between Ohio and the control widens monotonically, becoming statistically significant about a month after the lottery announcement on June 10.

A relatively small subset of COVID-19 cases require intensive hospital care and, reflecting the overall decline in Ohio's infection rates, we find that the vaccine incentive causes a decrease in ICU utilization as well (Figure 1.4). Panel (a) shows the cumulative total COVID-19-induced patient-days spent in hospital ICUs per 100,000 population by region and date. Panel (b) plots the difference between Ohio and its synthetic counterfactual. As discussed just above, because of the delay in immunity from the vaccine, any potential effect should not be expected until early June. Moreover, we should also expect some additional delay of a few days between disease onset and admission to a hospital ICU (Wang et al., 2020). Given this mechanical lag, it is unsurprising to see no effect on ICU patient-days in Ohio relative to counterfactual until the second week of June. However, ICU patient-days then also begin to decrease monotonically over time compared to the control, reaching statistical significance on June 13.[11]

Figure 1.5 shows the robustness of our estimated vaccination effects to different specifications. Similar plots for cases and ICU patient-days are provided in Appendix Figures A2 and A3. Each row of the figure(s) plots point estimates and 95 percent confidence intervals from a different model that varies either the specification, donor pool, or method of inference. Importantly, none of the model estimates have confidence intervals that include zero. The first row reproduces the estimates shown above from our baseline model. In the second row, we show results from the classic synthetic control

[11] To facilitate a more direct comparison of the timing and magnitude of these respective treatment effects, Appendix Figure A1 plots the estimates over time for all three outcomes during the post-treatment period.

model without covariates. As discussed in Section 1.2, the classic SCM produces a somewhat worse pre-treatment fit of the Synthetic Ohio—supporting our use of the ridge augmentation—but the estimates remain relatively close to the baseline values. The next row shows results using residuals of the outcomes to the state covariates, finding very similar results as the baseline model. In the fourth row, we estimate the baseline model but use Chernozhukov, Wüthrich, and Zhu's (2021) conformal inference method rather than the jackknife+ inference. Although the estimates become somewhat less precise—likely because conformal inference forces a sharp null hypothesis rather than using the leave-one-out pre-treatment residuals to proxy for the post-treatment variation under the null—the estimates remain statistically significant. Finally, we vary the state donor pool to assess sensitivity, both by including all 50 states and (separately) by iteratively leaving out each potential donor state.[12] On the whole, this extensive set of robustness exercises provides compelling support for the causal inference of our analysis.

The magnitudes of the estimated effects are shown in Table 1.3. In Panel [A], we detail the evolution of the lottery's effect on cumulative vaccination rates throughout the treatment period. One week into the intervention, we estimate that there is a 0.31 percentage points increase in vaccination rates in Ohio relative to the counterfactual. By the second week into the program, the estimated rate of increased uptake is 0.6 percentage points. The estimated effect then remains relatively stable over the remain-

[12] Appendix Figure A4 presents inference using the cross-sectional permutation approach of Abadie, Diamond, and Hainmueller, 2010 by estimating "placebo effects" for untreated units—comparing the ratio of the post-treatment root mean squared prediction error (RMSPE) and pre-treatment RMSPE of each of these estimates to that for the treated unit. We find strong support for Ohio's candidacy as an outlier among placebo-treated states.

ing three and a half weeks, reaching 0.7 percentage points by the end of the incentive program. Given that the counterfactual vaccination rate is 46.5 percent of population by the end of the treatment period, this end-line effect size corresponds to a 1.5 percent increase in the vaccinated share of Ohio's population. These effects shown in the table are all statistically significant, with the lower bound of the 95 percent confidence interval at 0.15 percentage points after the first week and at 0.52 percentage points by the end of treatment.

Panel [B] presents estimates for the lottery's effect on COVID-19 cases and ICU patient-days, both at the end of treatment (June 20) and at the end of the data four weeks later (July 18). We find statistically significant estimates for both of these outcomes by the time the treatment period ends, with the effect sizes increasing in magnitude until (at least) the end of the analysis period. By end of sample, Ohio has 125.3 fewer total recorded COVID-19 cases per 100,000 population (a 1.3 percent reduction relative to the counterfactual) and 40.56 fewer COVID-19-induced ICU patient-days per 100,000 population (2.5 percent of the counterfactual). At a glance, it may seem implausible that a 1.5 percent increase in vaccination rates could reduce COVID-19 outcomes by these magnitudes; however, this underscores the importance of accounting for the exponential nature of communicable disease transmission. Ohio's incentive program essentially serves as a shock, producing a surge in vaccinations in the state over a fairly short window of time. In turn, this reduction in the infection-vulnerable population abates transmission to unvaccinated people to yield further reductions in infections, effectively "bending the curve" for exponential growth of COVID-19 within the state. In addition, the type of person who is persuaded to vaccinate by a financial lottery-based incentive might be especially valuable in curtailing the spread of the disease.

Using these estimates from Table 1.3 along with data for Ohio, Table 1.4 shows calculated aggregate effect sizes and characteristics by time period. By aggregating the per-capita effects that we observe over the post-treatment period using state population, we are able to compute the total number of compliers, COVID-19 cases prevented, and ICU patient-days averted due to Ohio's intervention. The first column of the table shows calculated values for the latest 40 days of the pre-treatment period (April 2 through May 11, 2021), and the remaining three columns show various time windows during the (post-) treatment period. In the third column, we observe that about 690,000 Ohioans in total received their first/only vaccine dose during the 40 day span of the Vax-A-Million program. Using the estimates from Table 1.3, we calculate that 82,000 (12%) of these people did so only because of the treatment; i.e., they are treatment compliers.

Among the compliers, 86 percent chose to get vaccinated within the first two weeks of the lottery, thus making them eligible for all of the five prize drawings. Comparing first dose vaccinations across vaccine manufacturers, Janssen (Johnson & Johnson) retained about 12-13 percent market share both before and during the lottery initiative, while Moderna's market share in Ohio shrunk from 33 to 26 percent and Pfizer's grew from 55 to 62 percent. Appendix Figure A5 shows daily time series of manufacturers' market shares in Ohio and the United States during our study period. On the whole, there is little evidence that Ohio's treatment compliers systematically selected a different vaccine mix than the vaccinated population at large. The table also explores vaccination series follow-up. Although only first dose vaccinations were required for lottery eligibility, the evidence supports that compliers exhibit typical second dose follow-up rates. In the 40 days leading into treatment, 87 percent of people who start a Pfizer or Moderna vaccine series also obtain a second dose (using 21 and 28 day lagged windows, respectively, for counts of Pfizer and Moderna second dose vaccinations). During Ohio's program, this rate is 88 percent.[13]

Turning to the final column of Table 1.4, we find that Ohio's program substantially affects total COVID-19 cases and ICU utilization. By four weeks after the lottery's completion, we estimate that the program led to nearly 15,000 fewer cases and almost 5,000 fewer days spent in the ICU than would have occurred absent the lottery. Based on Di Fusco et al.'s (2021) values for COVID-19 ICU hospitalization, this amounts to about 325 fewer patients in the ICU for COVID-19-related complications. To reiterate, these aggregate effects only include reductions during our sample window, and the exponential nature of disease transmission implies that prevention of additional cases and ICU patients is likely as well.

[13] Follow-up rate is determined by summing total second doses of Pfizer and Moderna with 21 and 28 day lags, respectively, and then dividing this value by the sum of the total first doses of each. For all dates t within each period/column of the table, this formula is:

$$\text{Follow-up rate} = \frac{\sum_t \text{Pfizer(2nd dose)}_{t+21} + \text{Moderna(2nd dose)}_{t+28}}{\sum_t \text{Pfizer(1st dose)}_t + \text{Moderna(1st dose)}_t}$$

1.5 Conclusions

At the time of this writing, projections show that more than ten million people will have died from COVID-19 worldwide by October 2021, with additional deaths in following months (IHME, 2021). Of course, the realized extent of this death toll greatly depends on how many people are vaccinated. Safe and effective COVID-19 vaccines are freely available in many countries, including the United States, but this widespread vaccine availability is inadequate if a large portion of the population remains unwilling to vaccinate. A vaccine mandate could be used to increase vaccination rates (Abrevaya and Mulligan, 2011; Lawler, 2017), but making COVID-19 vaccinations mandatory is both publicly unpopular and politically tangled (Largent et al., 2020; Ivory, Leather, and Gebeloff, 2021). Ultimately, vaccination is a choice that depends on a person's beliefs about the benefits of being vaccinated (Auld, 2003).

To increase the perceived benefits of vaccination, a growing number of governments have implemented conditional cash lotteries (CCLs) that offer opportunities to win large prizes only available to vaccinated individuals. A CCL incentive is promising in this context because of its targeted nature: people with a greater propensity to decline vaccination are also more likely to assign a higher expected value to a lottery, a behavioral phenomenon known as probability neglect. Our paper evaluates the first CCL for COVID-19 vaccinations, which Ohio implemented during May and June of 2021. We find that Ohio's initiative significantly increases vaccinations—successfully encouraging more than 82,000 Ohioans who would otherwise not be vaccinated, an increase of

1.5 percent. Furthermore, we estimate that this surge in vaccinations then decreases COVID-19 prevalence within the state, reducing infections and ICU utilization by at least 1.3 percent and 2.5 percent, respectively.

These estimates allow us to assess the cost-effectiveness of the program. In a large study of COVID-19 patients, Di Fusco et al., 2021 find that the average hospital bill per day in the ICU is around 13,500 dollars. Using our estimate of the number of ICU patient-days averted, we calculate that the total benefit from avoiding these charges is approximately 65 million dollars. Additionally, there are substantial other social benefits from the 15,000 or more cases prevented, such as quality of life enrichment—especially for those who avoid cases of "long-haul COVID," where symptoms persist for months or longer—and potentially lives saved. Given that the total cost of Ohio's Vax-A-Million incentive scheme is 5.6 million dollars, the benefits of the CCL unquestionably exceed the program's cost.

Hesitancy towards vaccines has been rising globally in recent years, creating a significant challenge for policymakers. In lieu of mandates, governments are increasingly turning to other instruments to improve vaccination rates. Our evidence from Ohio's program illustrates that financial incentives—and conditional cash lotteries more specifically—are an effective means to increase vaccine uptake in areas plagued by vaccine hesitancy. Although a CCL is certainly not a panacea, we show that it can be a cost-effective component of a broader policy mix to increase vaccine uptake, with compelling potential to support other public health objectives as well.

References

Abadie, Alberto (2021). “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects”. In: *Journal of Economic Literature* 59.2, pp. 391–425.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program”. In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

Abadie, Alberto and Javier Gardeazabal (2003). “The Economic Costs of Conflict: A Case Study of the Basque Country”. In: *American Economic Review* 93.1, pp. 113–132.

Abrevaya, Jason and Karen Mulligan (2011). “Effectiveness of state-level vaccination mandates: Evidence from the varicella vaccine”. In: *Journal of Health Economics* 30.5, pp. 966–976.

Auld, M.Christopher (2003). “Choices, beliefs, and infectious disease dynamics”. In: *Journal of Health Economics* 22.3, pp. 361–377.

Barham, Tania and John A. Maluccio (2009). "Eradicating diseases: The effect of conditional cash transfers on vaccination coverage in rural Nicaragua". In: *Journal of Health Economics* 28.3, pp. 611–621.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2021). "The Augmented Synthetic Control Method". In: *Journal of the American Statistical Association* Forthcoming.

Biden, Joe (2021). "Remarks by President Biden Marking the 150 Millionth COVID-19 Vaccine Shot". The White House.

Björkman Nyqvist, Martina et al. (2018). "Incentivizing Safer Sexual Behavior: Evidence from a Lottery Experiment on HIV Prevention". In: *American Economic Journal: Applied Economics* 10.3, pp. 287–314.

Brehm, Margaret E., Paul A. Brehm, and Martin Saavedra (2021). "The Ohio Vaccine Lottery and Starting Vaccination Rates". Mimeo.

British Medical Journal (1957). "Vital Statistics". In: *The British Medical Journal* 1.5025, pp. 1013–1014.

Buchanan, Tyler (2021). "With first drawing days away, Ohio legislator seeks to stop Vax-A-Million lottery". In: *The Columbus Dispatch*.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu (2021). "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls". In: *Journal of the American Statistical Association* Forthcoming.

Dagan, Noa et al. (2021). “BNT162b2 mRNA COVID-19 Vaccine in a Nationwide Mass Vaccination Setting”. In: *New England Journal of Medicine* 384.15, pp. 1412–1423.

de Figueiredo, Alexandre et al. (2020). “Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: A large-scale retrospective temporal modelling study”. In: *The Lancet* 396.10255, pp. 898–908.

DeWine, Mike (2021). “Don’t Roll Your Eyes at Ohio’s Vaccine Lottery”. In: *The New York Times*.

Di Fusco, Manuela et al. (2021). “Health outcomes and economic burden of hospitalized COVID-19 patients in the United States”. In: *Journal of Medical Economics* 24.1, pp. 308–317.

Goette, Lorenz and Alois Stutzer (2020). “Blood donations and incentives: Evidence from a field experiment”. In: *Journal of Economic Behavior & Organization* 170, pp. 52–74.

IHME (2021). *GBD Compare: COVID-19 Projections*. Tech. rep. Institute for Health Metrics and Evaluation (IHME). University of Washington, Seattle, Washington.

Ivory, Danielle, Lauren Leather, and Robert Gebeloff (2021). “Least Vaccinated U.S. Counties Have Something in Common: Trump Voters”. In: *The New York Times*.

Lang, David, Lief Esbenshade, and Robb Willer (2021). “Did Ohio’s Vaccine Lottery Increase Vaccination Rates? A Pre-Registered, Synthetic Control Study”. Mimeo.

Largent, Emily A. et al. (2020). "U.S. Public Attitudes Toward COVID-19 Vaccine Mandates". In: *JAMA Network Open* 3.12, e2033324–e2033324.

Lawler, Emily C. (2017). "Effectiveness of vaccination recommendations versus mandates: Evidence from the hepatitis A vaccine". In: *Journal of Health Economics* 52, pp. 45–62.

Moran, William P et al. (1996). "Increasing influenza immunization among high-risk patients: Education or financial incentive?" In: *The American Journal of Medicine* 101.6, pp. 612–620.

Sunstein, Cass R. (2002). "Probability Neglect: Emotions, Worst Cases, and Law". In: *The Yale Law Journal* 112.1, pp. 61–107.

Thirumurthy, Harsha et al. (2016). "The Effects of Providing Fixed Compensation and Lottery-Based Rewards on Uptake of Medical Male Circumcision in Kenya: A Randomized Trial". In: *Journal of Acquired Immune Deficiency Syndromes* 72.Suppl 4, S299–305.

U.S. CMS (2021). *Biden-Harris administration increases Medicare payment for life-saving COVID-19 vaccine*. Tech. rep. U.S. Centers for Medicare & Medicaid Services.

Volpp, Kevin G., Leslie K. John, et al. (2008). "Financial Incentive–Based Approaches for Weight Loss: A Randomized Trial". In: *JAMA* 300.22, pp. 2631–2637.

Volpp, Kevin G., George Loewenstein, et al. (2008). "A test of financial incentives to improve warfarin adherence". In: *BMC Health Services Research* 8.272.

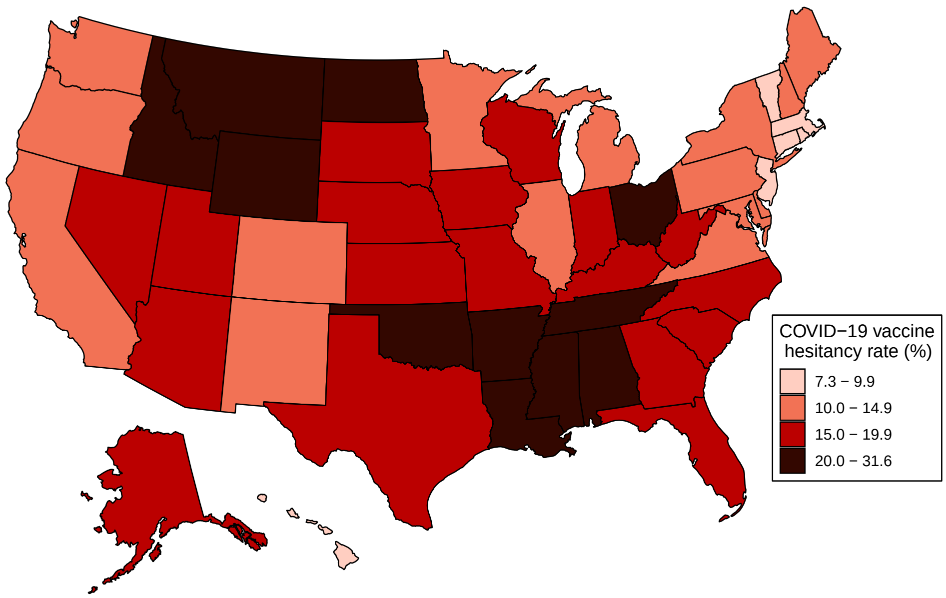
Vovk, Vladimir, Alex Gammerman, and Glenn Shafer (2005). *Algorithmic learning in a random world*. Springer.

Walkey, Allan J., Anica Law, and Nicholas A. Bosch (2021). "Lottery-Based Incentive in Ohio and COVID-19 Vaccination Rates". In: *JAMA* Forthcoming.

Wang, Dawei et al. (2020). "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China". In: *JAMA* 323.11, pp. 1061–1069.

1.7 Figures and tables

Figure 1.1: Surveyed COVID-19 vaccination hesitancy by state



Notes: Data plotted in this map use an average of the Census Bureau's Household Pulse Survey responses during Weeks 25-29 (February 17 to May 10, 2021). Vaccination hesitancy includes responses of "definitely not" and "probably not" as survey respondents' stated willingness to be vaccinated for COVID-19.

Table 1.1: Largest five unit weights for Synthetic Ohio using ridge augmented synthetic control for an outcome of the share of state population with any COVID-19 vaccination

State	Unit weight
Wisconsin	0.321
Kansas	0.281
Michigan	0.191
Idaho	0.181
North Dakota	0.126

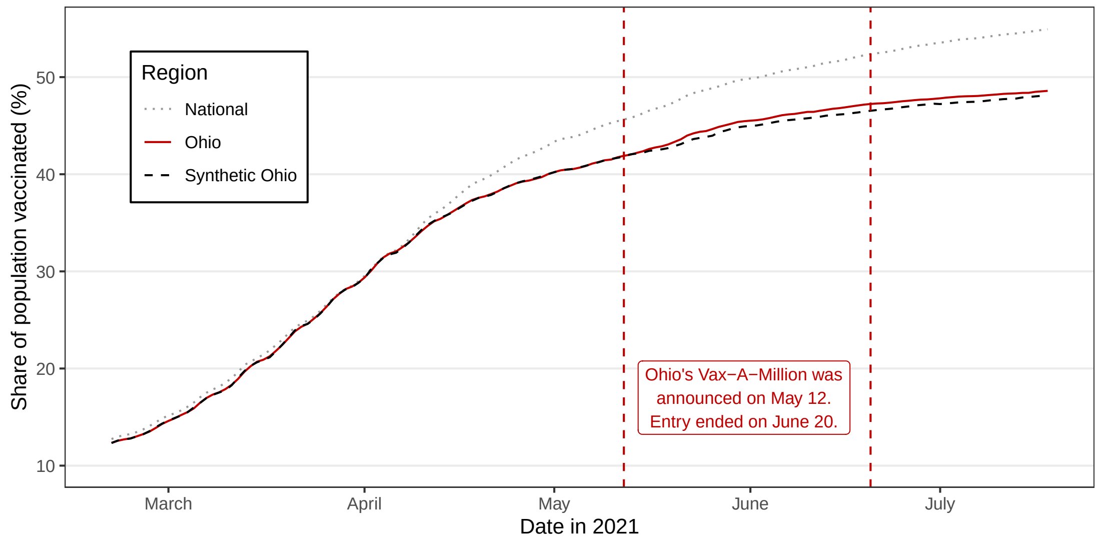
Notes: Online Appendix Table A2 shows the full set of unit weights for Synthetic Ohio.

Table 1.2: Summary statistics for the United States, Ohio, and Synthetic Ohio

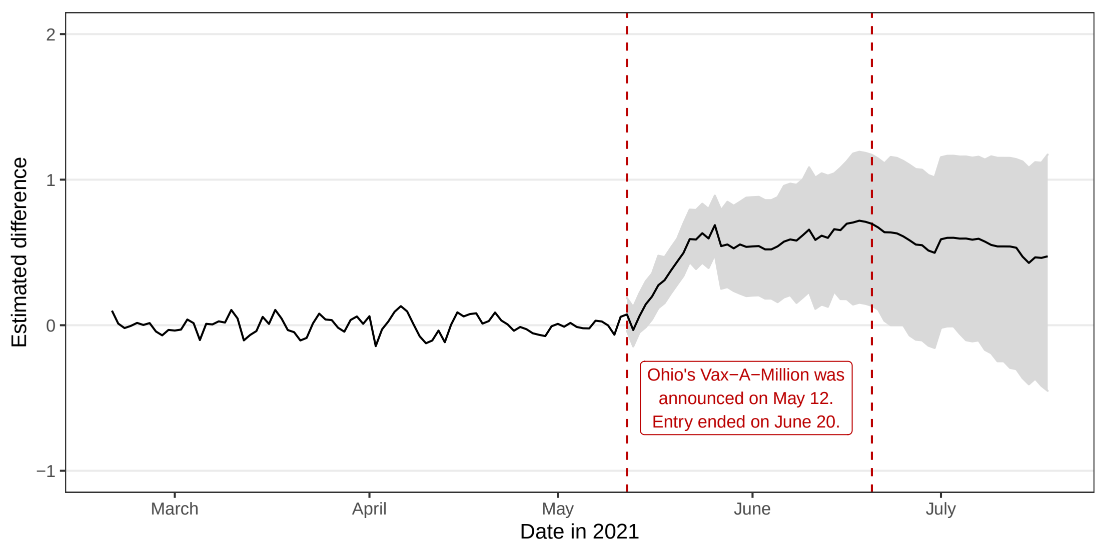
	State avg.	Ohio	Syn. Ohio
Panel [A] Outcome vars in pre-treatment period			
Share of pop. with any vaccination by April 2, 2021 (%)	30.21	30.04	30.18
Share of pop. with any vaccination by May 12, 2021 (%)	45.63	41.90	41.82
COVID-19 cases per 100k pop. by May 12, 2021	9,576	9,214	9,213
COVID-19 ICU patient-days per 100k pop. by May 12	1,430	1,450	1,449
Panel [B] Covariates in pre-treatment period			
Share of population of age 12 to 17 (%)	7.47	7.54	8.16
Share of population of age 18 or older (%)	77.06	77.22	74.46
Population density in 2020 (people per square mile)	423.64	288.80	240.56
Gross domestic product per capita in 2020 ($)	61,791	57,209	57,905
Republican presidential vote share in 2020 (%)	49.12	53.27	53.53
Influenza vaccination rate in 2019 (%)	47.41	50.00	49.99
Community Mobility Report for retail/recreation	-7.66	-5.59	-5.63
Community Mobility Report for grocery/pharmacy	-1.22	-1.65	-1.56
Community Mobility Report for parks	31.04	70.58	70.16
Community Mobility Report for transit stations	-15.77	-8.59	-8.41
Community Mobility Report for workplaces	-23.30	-20.99	-21.02
Community Mobility Report for residences	6.13	5.30	5.32
Med. dist. of pop. to closest vaccination site (mi)	1.10	0.91	0.91
95th percentile dist. to closest vaccination site (mi)	11.08	7.01	6.99

Notes: Table 1.2 presents sum. stats for the U.S, Ohio, and Syn. Ohio. Ohio's Vax-A-Million incentive program was announced on May 12, 2021 and lottery entry ended on June 20, 2021. Panel [A] shows values for the dep. variables during the pre-treatment time period(s) indicated. These outcomes are: the share of pop. with at least a first dose of any COVID-19 vaccination, the cumulative total COVID-19 cases per 100k pop., and the cumulative total COVID-19 hospital ICU patient-days per 100k pop.. Panel [B] shows values for state covariates during 2019 or 2020, as indicated, or during the pre-treatment analysis period of February 19 through May 11, 2021.

Figure 1.2: Share of population with any COVID-19 vaccination over time



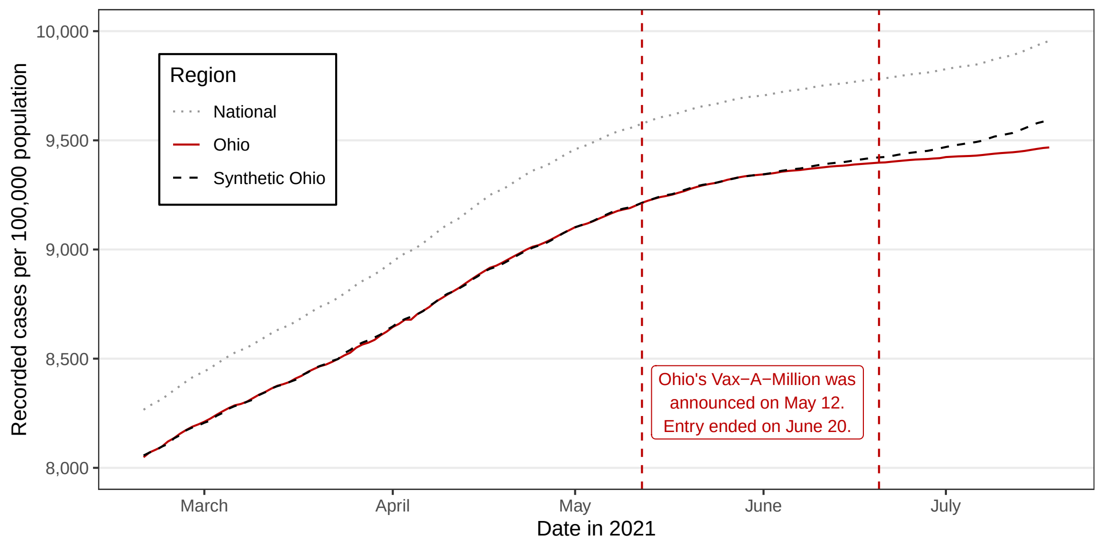
(a) Time series plotted by region



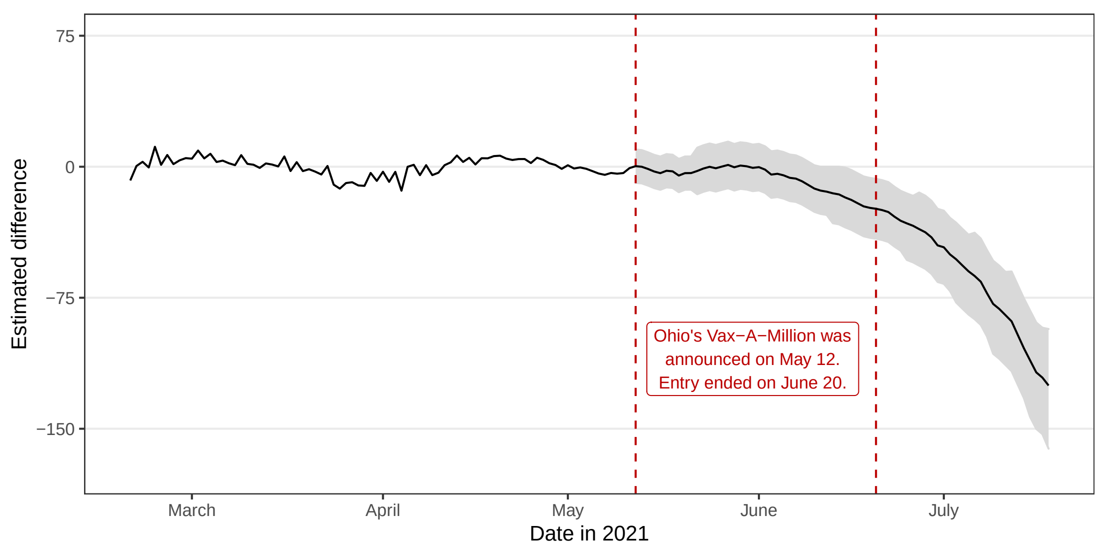
(b) Estimated difference between Ohio and Synthetic Ohio

Notes: Panel (a) of Figure 1.2 shows time series graphs for the share of population that had received at least a first dose of any COVID-19 vaccination by region and date. Panel (b) shows the estimated difference between Ohio and the synthetic control. The grey shading indicates 95 percent confidence intervals for each post-treatment date, calculated using conformal inference.

Figure 1.3: Cumulative total COVID-19 cases recorded per 100,000 population over time



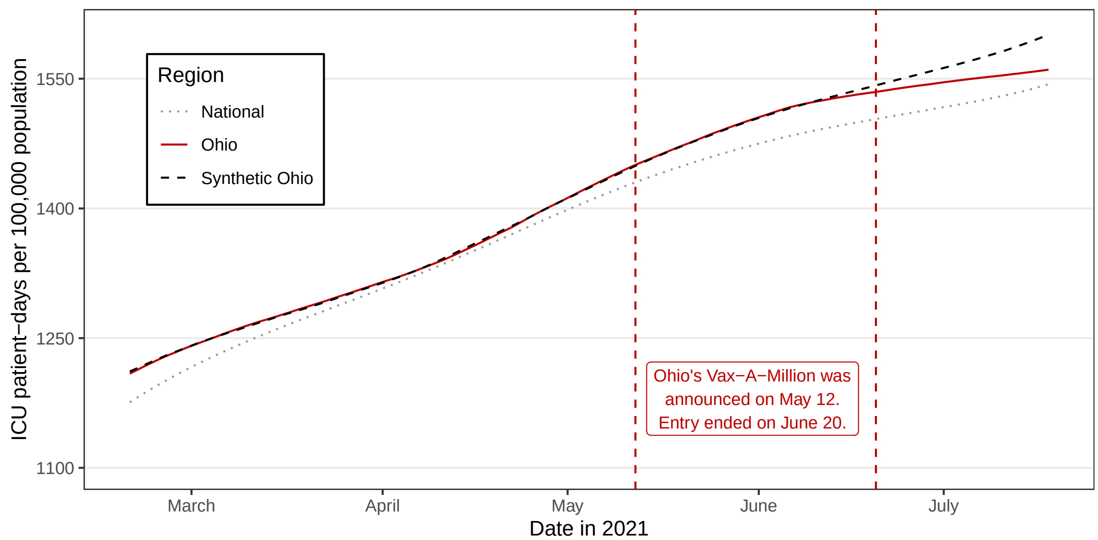
(a) Time series plotted by region



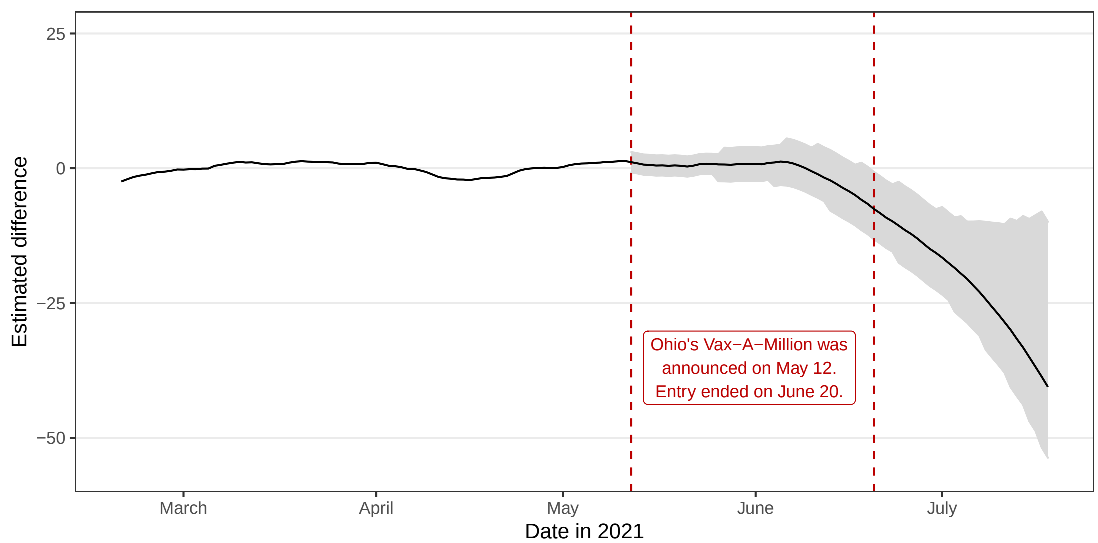
(b) Estimated difference between Ohio and Synthetic Ohio

Notes: Panel (a) of Figure 1.3 shows time series graphs for the cumulative total number of COVID-19 cases (positive COVID-19 tests) recorded per 100,000 population by region and date. Panel (b) shows the estimated difference between Ohio and the synthetic control. The grey shading indicates 95 percent confidence intervals for each post-treatment date, calculated using conformal inference.

Figure 1.4: Cumulative total COVID-19 ICU patient-days per 100,000 population over time



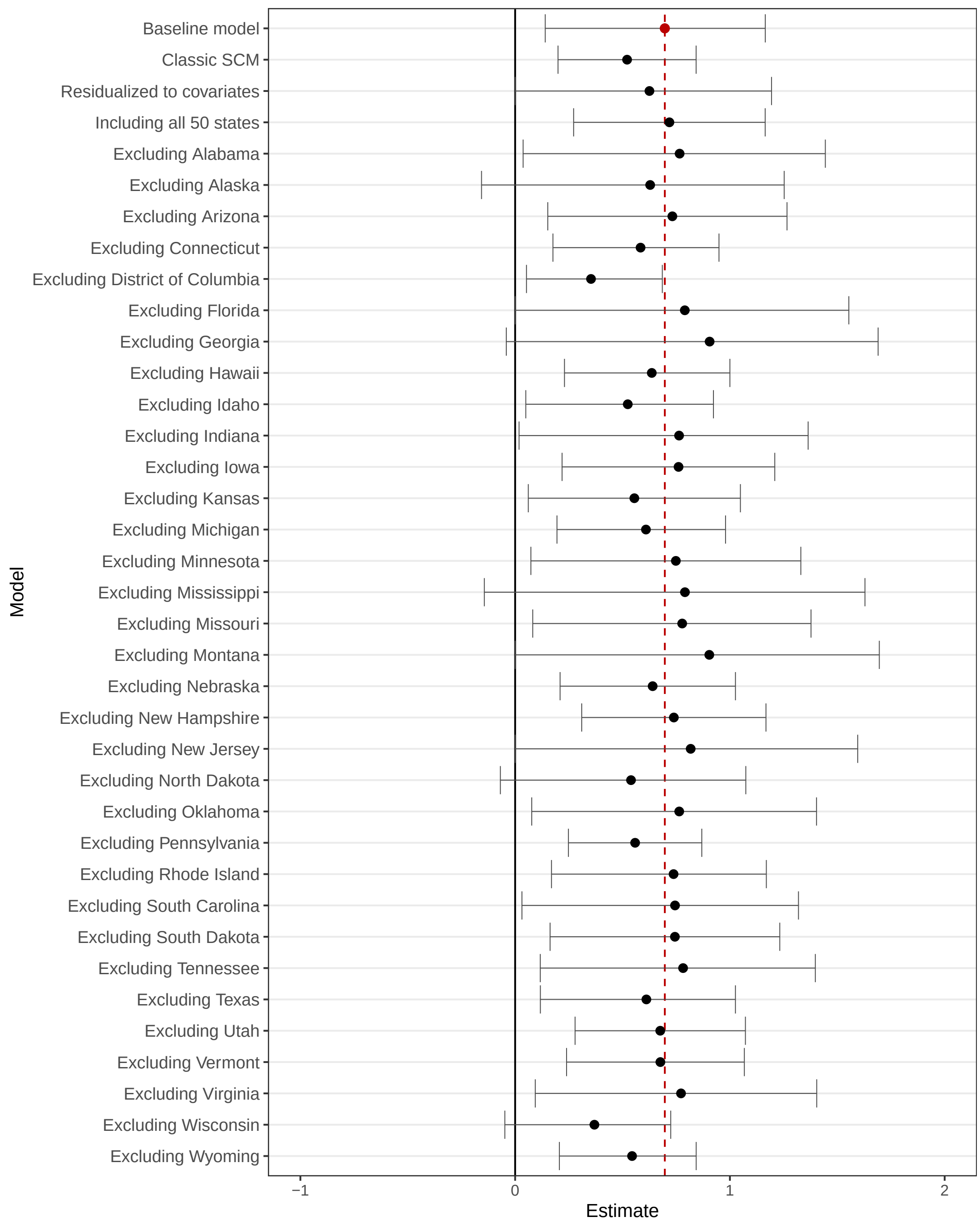
(a) Time series plotted by region



(b) Estimated difference between Ohio and Synthetic Ohio

Notes: Panel (a) of Figure 1.4 shows time series graphs for the cumulative total COVID-19 hospital ICU patient-days per 100,000 population by region and date. Panel (b) shows the estimated difference between Ohio and the synthetic control. The grey shading indicates 95 percent confidence intervals for each post-treatment date, calculated using conformal inference.

Figure 1.5: Robustness checks of SCM estimates for the share of population with any COVID-19 vaccination by the end date, using diff. samples and specifications



Notes: Figure 1.5 shows estimated differences between Ohio and the synthetic control for the share of population that had received at least a first dose of any COVID-19 vaccination by June 20, 2021. Each row depicts results from a separate model using the data sample and/or specification denoted. Grey error bars indicate 95 percent confidence intervals, which are calculated using conformal inference.

Table 1.3: Estimation results for Ohio compared to the synthetic control

Outcome	Date	Estimate	95 pct. conf. interval		Cf. value
			Low bd.	Up. bd.	
Panel [A] COVID-19 vaccinations during lottery treatment					
Population vaccinated (%)	May 18	0.3098	0.1520	0.4676	42.6
Population vaccinated (%)	May 25	0.5959	0.3930	0.7988	43.9
Population vaccinated (%)	June 01	0.5415	0.2034	0.8797	45.0
Population vaccinated (%)	June 08	0.5815	0.1531	0.9647	45.6
Population vaccinated (%)	June 15	0.6531	0.1797	1.081	46.1
Population vaccinated (%)	June 20	0.6970	0.1334	1.170	46.5
Pop. vaccinated 18-older (%)	June 20	0.7761	0.2953	1.257	57.8
Panel [B] COVID-19 infections during and post-treatment					
Cases per 100k population	June 20	-24.06	-41.19	-6.932	9,422
Cases per 100k population	July 18	-125.3	-161.4	-92.90	9,593
ICU patient-days per 100k pop.	June 20	-7.540	-13.17	-0.6617	1,542
ICU patient-days per 100k pop.	July 18	-40.56	-53.70	-9.927	1,601

Notes: Table 1.3 shows results from ridge augmented synthetic control estimations for Ohio's Vax-A-Million incentive program, which was announced on May 12, 2021. Lottery entry ended on June 20, 2021. The outcomes in rows are the share of population with any COVID-19 vaccination, the cumulative total number of COVID-19 cases recorded per 100,000 population, and the cumulative total COVID-19 hospital ICU patient-days per 100,000 population. The 95 percent confidence intervals are calculated using conformal inference. The final column shows the counterfactual values from Synthetic Ohio.

Table 1.4: Aggregate estimated effects and characteristics for Ohio by time period

		Vax-A-Million treatment period		
	Pre-treatment	First two weeks	Full period	Post-treatment
Date range included	April 2 - May 11	May 12 - May 25	May 12 - June 20	May 12 - July 18
Number of days	40	14	40	68
Ohio population	11,799,448	11,799,448	11,799,448	11,799,448
Vax-A-Million cost		5,600,000	5,600,000	5,600,000
Total first dose vaccinations	1,488,978	339,226	690,135	
First dose compliers		70,315 (21%)	82,239 (12%)	
First dose always-takers		268,911 (79%)	607,896 (88%)	
First dose of Janssen	174,651 (12%)	40,186 (12%)	87,648 (13%)	
First dose of Moderna	491,329 (33%)	92,300 (27%)	176,943 (26%)	
First dose of Pfizer	822,998 (55%)	206,740 (61%)	425,544 (62%)	
Program cost per complier		80	68	
2nd dose Moderna in 28 days	403,991	80,568	142,786	
Moderna follow-up rate	82%	87%	81%	
2nd dose Pfizer in 21 days	736,942	168,624	388,493	
Pfizer follow-up rate	90%	82%	91%	
Overall follow-up rate	87%	83%	88%	
COVID-19 cases prevented				14,779
ICU patient-days prevented				4,786

Notes: Table 1.4 uses data for Ohio and estimates from Table 1.3 to calculate aggregate effect sizes and characteristics for Ohio by time period. All dates included are in 2021. The row for second doses of Moderna uses a time period shifted forward by 28 days, e.g., using second doses during June 9 - July 18 for the full treatment period of May 12 - June 20 column. Similarly, the Pfizer second dose row uses a time period shifted by 21 days. The second dose follow-up rates are calculated by dividing the total Moderna and/or Pfizer second dose values by the total Moderna and/or Pfizer first dose values.

Chapter 2

Too Legit to Quit? Labor Supply Responses to the Four-Day School Week

2.1 Introduction

> We have a moral and ethical obligation to make [the four-day week] happen... Do not let nostalgia for antiquated systems hinder or prevent improvement and system change for today's generation of children.
>
> — Donald Kordosky, *The Four-Day School Week: Less is More!*, 2011

> This is something that's happening, nobody's really evaluating it, nobody's asking what should be the minimum required if somebody's going to do it. The states are just letting it happen, and it's unfortunately going to be very hard to reverse because it's one of those adult-benefit things that you can't roll back.
>
> — Paul Hill, *The Atlantic*, March 2017

Recruitment and retention of qualified schoolteachers are some of the primary challenges currently facing many schools today. While this phenomenon can be at least partially attributed to the COVID-19 pandemic and the toll it has taken, it is likely the collision of this short-term fluctuation with a longer-term trend of declining participation in the teaching workforce. However, for all of the discussion in the national media about a looming "teacher shortage" in the United States, it's not clear that the problem is one affecting all states equally, and this is evidenced by some of the extraordinary measures some states have taken to keep existing teachers, recruit (or poach) credentialed teachers, or even permit teachers that would have previously been considered unqualified.

The difficulty in answering the teacher shortage question is exacerbated by the lack of data, especially at the national level (Ingersoll, 2003; Garcia and Weiss, 2019). In fact, at present, neither the federal government nor the majority of states collects and publishes data detailing the intensity of teacher shortages being experienced in each state (Walsh, 2016). Some states monitor their existing teaching force, but not all forecast future teacher demand based on expected enrollments, approaching retirements, or how many positions are left unfilled, and such information, if available, is minimal and usually not tracked over time (Learning Policy Institute, 2022). Thankfully, there have been recent contributions that have come from outside of education departments that, for the first time, shine some light on the magnitude and location of teacher shortages around the country (Sutcher, Darling-Hammond, and Carver-Thomas, 2019; Nguyen, Lam, and Bruno, 2022). The U.S. Department of Education does collect and report the subject areas in which states report teacher shortages (TSAs). However, they do not indicate the magnitude of shortages in each subject, and recent years' reports simply list nearly every subject taught in states experiencing shortages (Cross, 2017).[1] Now more than ever before, given the toll the pandemic has taken on educators, the urgency in addressing teacher shortages has reached a crescendo, as it is becoming more difficult for districts to find candidates to fill vacancies – at a time where students need teachers more than ever after experiencing never-before-seen learning losses throughout the country (Hanushek and Woessmann, 2020).

Data limitations notwithstanding, recent research has shown that teacher shortages appear to be more regional than originally thought, with Southern and Midwestern states bearing the brunt of these pressures (Nguyen, Lam, and Bruno, 2022). However, there are at least two reasons why shortages are still being observed. First, from a practical standpoint, the supply of teachers is defined as the pool of "qualified individuals willing to offer their services under prevailing wages and conditions" (Sutcher, Darling-Hammond, and Carver-Thomas, 2019). Because teacher wages have fallen both in real terms and relative to other college graduates in many parts of the country, the supply of teachers willing to work in certain states is often much smaller than the raw number of individuals qualified to teach. These effects can be seen upstream in the teacher pipeline – the number of college graduates with a bachelor's degree in education has decreased 19 percent since 2000-2001, and this decrease is even more pronounced when going back to 1970 – a year that produced twice as many education graduates

[1] The primary reason for these reports seems to be that teaching in one of these teacher shortage areas makes one eligible for debt forgiveness – a tool that the current administration has stated that it would like to expand and one that has been used for teachers in North Carolina and Florida and for doctors who commit to work in underserved areas.

than we had in the 2019-2020 graduating class (Schaeffer, 2022).

Second, the supply and demand for teachers is a challenge to characterize because there is simply not a national labor market for teachers. Such an absence has led to a recent emphasis on “grow your own” teacher pipeline programs, whereby communities start developing future teachers while they are still in the local school system.

Developing future teachers while they are still in middle and high school is only one of many strategies that teacher-strapped states and districts have used to tackle these shortages. A variety of other approaches have been used to develop, attract, and/or keep teachers, most notably increasing pay. In 2018, seven states saw state-wide teacher strikes that led to pay increases, but not all strikes were successful and both states and districts are often extremely limited in their ability to offer raises, so states have had to get creative. In addition to pay increases, which come with compounding costs, schools have used one-off bonuses, loan forgiveness, education cost-sharing, and housing subsidies.

Alternatively, states with less financial flexibility have attempted to remove or reduce barriers to entry with programs that reduce the costs (or increase the benefits) of serving as a teacher. Perhaps the most radical of these approaches has been the introduction of the four-day school week. The use of a non-monetary amenity like the four-day school week provides an opportunity to study the effect on the labor market for teachers across two separate dimensions: retention and recruitment. With regard to recruitment, we have anecdotal evidence that the four-day school week has served as an effective recruitment tool for schools that have struggled to hire qualified teachers in recent years. The power of such a schedule change, in lieu of a pay raise, has also been seen in the private sector as well. Many employers have either felt pressure to shift to a four-day work week and/or a hybrid model that combines the pandemic-driven work-from-home paradigm shift with less-frequent traditional days in the office.

With regard to the effect such changes have on employee/teacher retention, this is an unanswered question. Given the fact that there seems to be a revealed preference for shorter and/or more flexible work weeks, the switch to a four-day week could be considered equivalent to a pay raise, we might expect teachers to attrit at lower rates using an efficiency wage or compensating differential framework (Stiglitz, 1976; Schlicht, 1978; Salop, 1979; Rosen, 1986).

However, merely retaining existing teachers and/or recruiting new ones are not the only objectives that a school needs to worry about. Teacher quality, a sub-dimension of both, is a vital concern for school districts since research shows that it is the most important schooling factor influencing student achievement (Goldhaber and Hansen, 2010). It is for this reason that we should be cognizant of potential heterogeneity in response to (perceived) pay increases. While one might assume that pay raises improve the quality of teachers, prior evidence suggests that higher teacher salaries might differentially retain lower quality teachers (Clotfelter, Ladd, and Vigdor, 2011). Hendricks (2014) provides some nuance to this analysis by proposing that teacher pay raises have the effect of retaining low-ability teachers in the first two years but also retain high-ability teachers who are tenured (typically three or more years of experience). Of course, the effect of pay raises is far from settled, just like the value-added literature that it naturally intersects with, but neither branch has taken up investigating the effects of the four-day school week to date. It is precisely this area where there is room for inquiry into how teachers respond to the implementation of a compressed work week, generally perceived as a non-monetary amenity.

In an effort to shine some light on this area, I estimate the impact of the four-day school week on the teacher labor market using an assembled panel of employment records for certified public school teachers in Oklahoma. Oklahoma is a prime candidate for this analysis because it has been a rapid adopter; 12% of schools have begun using the alternative schedule in only the last twelve years.[2] In addition, I use the staggered rollout of the policy between 2010 and 2019 to identify the four-day week's effect on teachers using an event study design. I find evidence suggestive of a reduction in attrition, and that this effect is primarily driven by new teachers who have selected into the schedule change. In addition, I find multiple pieces of evidence relating to the effect of four-day school week adoption. First, schools see a gradual increase in average experience after the schedule switch. This is at least partially driven by new teachers arriving with more experience than they did pre-policy; i.e., these schools are better able to compete for experienced teachers on the labor market. Furthermore, schools that adopt see a gradual reduction in their need for emergency credentials, which are seen as a last resort for districts that are unable to hire qualified teachers. While these results don't

[2] 207 of 1,857 schools in 2017-2018. Some of these schools have since switched back to five days in the years since.

eliminate the possibility of negative selection as a result of the policy, when combined with the null results on student achievement of Morton (2021), they certainly suggest that it is unlikely.

The remainder of the paper is organized as follows: Section 2 provides background information about four-day school weeks; Section 3 describes the dataset that was assembled; Section 4 lays out the empirical strategy used; Section 5 discusses the results; Section 6 concludes.

2.2 Background

2.2.1 Where and Why

This fall, students from nearly 2000 schools across 24 states are attending school for only four days each school week 2.1. This unconventional scheduling policy has existed for many years in rural school districts throughout the Intermountain West, but its popularity has increased quite rapidly in recent years, leading to numerous other states around the country adopting the alternative schedule 2.2. One particularly notable example is Oklahoma, which has had 222 of its 1,784 schools implement four-day school weeks since 2009 – the year in which schools were first permitted to make the change 2.3.[3]. Moreover, the majority of these schools have changed schedules since 2015. Similar uptake has been seen in Arizona, New Mexico, Montana, and most

[3]Four-day school weeks first became possible in Oklahoma with the passing of House Bill 1864 in the wake of the Great Recession in April 2009 (H.B. 1864 [2009]) The policy changed the requirements for districts such that they no longer needed to operate for 180 days and 1,080 hours of classroom instruction per year; rather, they only needed to meet the 1,080 hours requirement. As a result, districts could operate for fewer than 180 days per year if they met the required 1,080 hours.

recently, Missouri.

Until only the last few years, the impetus for adopting the schedule change has almost always been financial, as evidenced by the increase in take-up after the 2008 financial crisis (and the 1973 oil embargo before that), with savings coming, quite mechanically, from reduced transportation and overhead costs. Because of their geographic size and the amount spent busing students, larger, rural districts are tempted by the theoretical twenty percent reduction in transportation costs (going from five days to four). To be sure, these costs are not trivial, either. The amount spent on bus operations can exceed five percent of a district's total budget in some areas, and administrators have referred to the realized savings from reduced fuel usage in terms of "number of teachers [salaries] saved".[4]

In addition, all schools stand to realize savings on overhead costs from non-operation one day per week, regardless of transportation needs. This comes from a decrease in spending on food, utilities, administrative services, and non-salaried staff, e.g., custodians and cafeteria workers. However, savings do not (and cannot) come from a reduction in teacher salaries, which remain fixed. For cash-strapped districts, the temptation to compress the school week into four days to save money seems like the only way to make ends meet. Unfortunately, the savings are rarely what they are predicted to be – often ranging from 2 to 5% of total expenditures, with the majority

[4] kfor.com/2017/02/07/four-day-school-week-paying-off-for-local-districts-lawmakers-want-a-change/

of cases falling in the bottom of (if not below) that range.[5], [6]

So, if schools aren't actually saving money making this switch, then why do schools keep adopting the policy en masse and why do so few ever change back?

Four-day school week advocate, Donald Kordosky, argues that, "Most districts originally look at the four-day week as a money-saving tactic. If that is the only reason they are missing the boat." Korodosky considers the effect of four-day weeks on student achievement, schedule consistency throughout the year, attendance rates for both students and teachers, discipline, homework, parental effects, and teacher training/preparation. Certainly, there are many factors to be weighed by districts that are considering maintaining or adopting such a potentially repercussive policy. However, despite the rapid and continuing adoption of four-day school weeks, there is has been relatively little research on the impacts of the schedule change, and the rate of adoption only continues to outpace research into its effects.

Student achievement should, of course, be a primary concern when considering the adoption of this policy, but how do teachers fare under the altered schedule? Some accounts indicate that the time savings has led to more effective teaching and more attentive students. In fact, the numerous pieces written on the four-day week generally support the notion that student achievement should *theoretically* not be adversely affected by the alternative schedule and could potentially be beneficial (Kordosky, 2011; Tharp, 2014). The mechanisms and causal channels behind this speculated positive

[5] www.ecs.org/clearinghouse/93/69/9369.pdf

[6] sde.ok.gov/sde/newsblog/2017-04-13/study-minimal-cost-savings-4-day-school-weeks-increases-majority-districts

effect on student test scores are numerous, but this paper focuses on the way in which teachers' labor supply is affected. If the schedule change were to alter attrition patterns at treated schools, such compositional changes could impact student test scores, but this is merely conjecture and outside the scope of this paper. Similarly, four-day school weeks may alter the workforce by way of recruitment. Anecdotally, districts have begun to claim recruitment and retention as justifications, albeit generally ex-post, with the notable exception of Missouri which will be addressed later.

Unfortunately, recent evidence paints a very different picture of the effect that the compressed school week has on student achievement. Paul Thompson, the premier researcher in four-day school week research, finds that removing one school day each week is associated with a reduction in test scores in both reading and math (Paul N Thompson, 2021). However, he argues that this is largely due to a reduction in instructional time, so this could be mitigated in schools that manage to avoid such a decrease (Paul N. Thompson and Ward, 2022). His work largely sidesteps looking at effects on teachers, so it is agnostic on the competing effects that changes in teacher composition could have on this ; i.e., effects might be mitigated by positive selection or exacerbated by negative selection.

2.2.2 History of the Four-Day School Week

The traditional school year in America's public schools is approximately 180 days, which typically follows an agrarian calendar. Many in the educational field believe that the conventional school year was established to meet the needs of the 19th-century farmer who would undoubtedly need his children during harvest season while others have posited that the school calendar in the United States was established to allow students to vacate the urban areas during the heat of the summer (Pedersen, 2012). Regardless of the motivation, the calendar in the American education system is quite consistent across all 50 states. Despite recent efforts from educational groups to add more hours to the school day and/or add days to the school year, this traditional format of a five-day week for 180 days is not likely to change. In fact, to the chagrin of many in the field of education, a growing number of school districts have decided to go the other direction: **reducing** the number of school days.

The four-day week is far from new, however. Surprisingly, it's existed as a concept for nearly 100 years. First used by a school in Madison, South Dakota in 1931, the non-traditional schedule had a resurgence during the energy crisis of the 1970s as a solution to the escalating transportation costs faced by larger rural districts (Hunt, 1936; Johnson, 1977). In the years since, several states have revised their laws regarding the mandatory minimum number of school days (between 175 and 186) and replaced them with hours requirements instead. The minimum amount of instruction time varies by grade level, but all states fall between 900 (Alaska, Idaho) and 1137 (Wisconsin) hours (Rowland, 2014).

Since the choice to switch from five days to four occurs at the local (typically district) level, schools have a great deal of flexibility in how they choose to meet these hours requirements and details vary from district to district. In fact, rather than evaluating the pros and cons of adoption, much of the education literature on this topic discusses the various schedule choices schools can make in their implementation. This would typically entail choosing the day to be dropped (usually Monday or Friday) and how to spread this time across the four days that are retained, i.e., starting earlier and/or staying later. Generally speaking, under the four-day system, school days only need to be lengthened by 45-60 minutes to equalize instructional time between the two schedules and the duration of the school year remains unchanged. [7]

2.2.3 How Might the Four-Day Week Affect Teachers?

There are a number of potential mechanisms for a compressed weekly schedule to affect teachers, which might at least partially explain why there is still no clear consensus on whether or not schools are worse off under the four-day regime, all things considered.

First, consider the response of teachers to this modification of their work schedule. Some have argued that longer class periods give teachers the flexibility to organize lessons more effectively and utilize a variety of different teaching methods (Rice, Croninger, and Roellke, 2002). In fact, surveys of teachers at treated schools reveal that many believe less time is wasted during the four-day week, leaving more time for instruction. Furthermore, policy supporters argue that teachers are able to better manage their time because their instruction is more focused and longer class periods improve curriculum continuity.

[7] Paul N. Thompson and Ward (2022) finds that many schools are actually *not* maintaining instructional hours through this schedule change with treated students receiving approximately 85 fewer instructional hours, on average, over the course of a school year

In addition, some districts use the day off to give teachers time to plan lessons and collaborate with other faculty. The difficulty presented by these hypothesized channels, though, is that they are quite difficult to test. However, one possible effect that is testable is reduced turnover and absenteeism. Higher teacher turnover and frequent absenteeism have both been shown to negatively affect student achievement, and anecdotal evidence suggests that both might be reduced under the alternative schedule (Miller, Murnane, and Willett, 2008; Ronfeldt, Loeb, and Wyckoff, 2013).[8]

Also, it could simply be the case that teachers prefer working four days instead of five and that any productivity gains are due to improved morale. This is certainly consistent with both survey data on teachers in four-day districts and the limited research on four-day work weeks (Turner, Finch, and Uribe-Zarain, 2018; Hamermesh and Biddle, 2022). While most of the potential impacts for teachers appear to be positive, there is still a cost associated with the transition; current teachers are forced to redevelop curriculum to better fit the longer class periods, so teaching quality may suffer in the short-term as teachers sort across school types and adjust to a different schedule.

Finally, it is important to be aware of possible selection effects, which could potentially be playing a role here for teachers at the margin. Arguments for the existence of positive or negative selection could both easily be made. On the one hand, if we believe that teachers perceive this to be an amenity that has utility value, this could induce (a) good teachers with experience who might otherwise leave for better pay elsewhere to

[8] While efforts were made to get data on teacher absences and usage of substitute teachers, OSDE was not able to provide data on either.

stay or (b) experienced teachers who are teaching elsewhere (or who aren't teaching) to apply to schools that offer it. On the other hand, if we believe that offering three-day weekends every week makes the job more appealing, this could induce teachers who would've previously selected out of teaching (revealing themselves to be less committed teachers) to consider teaching under four-day conditions.

2.3 Data

I've compiled a panel dataset that draws on four sources: school data from the National Center for Education Statistics Common Core of Data (NCES CCD), employment records for teachers in Oklahoma from the Oklahoma State Department of Education (OSDE), county-level economics data, and an author-assembled panel of schools that switched schedules during the analysis period. In addition, I also assembled a list of districts that were forced to use emergency credentials (explained below) during the same years spanned by the previous dataset.

2.3.1 OSDE Employment Records

The OSDE compiles an annual report of all teachers and staff in the public school system, posted to their website every October.[9] The reports include the universe of public school teachers in Oklahoma between the 2006-2007 school year and the 2021-2022 school year. These records provide each teacher's first and last names; total teaching experience; county, district, and school taught in; race and gender; grade(s)

[9] sde.ok.gov/documents/2018-01-02/certified-staff-salary-information

and subject area(s) taught; position title and salary. Unfortunately, the system for identifying teachers was changed twice during the analysis period, so teachers are not always linked across years. Great care was taken to longitudinally link the three systems using author-written sequential matching algorithms to track name inconsistencies, name changes due to marriage and, and school changes within district. Furthermore, it appears that there are also data quality issues with the teacher identification for teachers who have multiple teaching spells during the analysis period with different ID numbers being issued for each spell, so these individuals were nested within a new identification number.

2.3.2 National Center for Education Statistics CCD

Next, I use the National Center for Education Statistics (NCES) annual Common Core of Data (CCD) for demographic information about each school. This data set includes the racial makeup of each school, enrollment counts, percentage of students receiving free lunch, and number of full-time equivalent employees (FTEs) in addition to information about school location (address, latitude/longitude, and locale type). These will provide some of the school-level controls in my regression specifications. In addition, I test to see if any of these variables are changing discretely at the time of adoption. These tests are included in the Appendix.

2.3.3 Four-Day School Week Panel

The OSDE began producing an annual list of four-day schools operating within the state starting with the 2010-2011 school year. This list has been maintained until 2018-2019 where it was discontinued because of the pandemic. Since four-day school weeks began before the first year in this panel and continued after 2019, missing data has been gathered in joint work with Paul Thompson to complete the panel. Treatment status for the schools is reflected in a panel describing the year-by-year uptake of the four-day school week for Oklahoma schools, which begins in the 2009-2010 school year and accelerates every year until 2017 at which point several schools begin switching back.

2.3.4 ACS 5-year Estimates

In order to capture local economic conditions, I use the American Community Survey (ACS) from the US Census Bureau.

2.3.5 Emergency Credentials

In the event that a school is unable to hire a qualified teacher, Oklahoma law provides that the State Board of Education may issue an emergency certificate, as needed. Despite being a law on the books since 1997, it was not used for the first time until 2011. The OSDE publishes an annual list of districts that were approved, along with the both the number of certificates and number of subject areas they were issued for. The growth of these credentials can be seen in Figure 2.7. These are used as an outcome in my event study design since the need for them should be negatively correlated with both teacher quality and ease of hiring.

2.4 Empirical Strategy

2.4.1 Event Study

Our primary identification strategy is the use of an event study design that exploits the staggered adoption of the treatment. Naturally, this greatly restricts our sample, but since these schools are quite idiosyncratic, schools that eventually become four-day schools likely serve as the best controls for the earliest adopters. It is important to state that this design imposes the crucial assumption that, conditional on being treated, the exact timing of treatment is **as good as random**. This assumption seems reasonable in this setting since we (a) cannot find reliable predictors of treatment and (b) early adopters aren't very different than late adopters with the exception of smaller enrollments.[10] It should also be noted that this assumption has been made in all of the extant four-day school week effect literature. Given this assumption, the key to measuring effects with the event study design is centering policy adoption in *event time*. This is done by subtracting the year of adoption from the year of observation for all treated units. The event time variable can then be translated into indicator variables and used in the style of Equation 2.1 below.

[10] These differences are small and driven by the existence of a few particularly small schools.

$$y_{i,s,d,c,t} = \delta_t + \delta_s + \sum_i \beta_i \text{Yrs before FDSW}_{s,t} + \sum_j \beta_j \text{Yrs after FDSW}_{s,t} + \epsilon_{i,s,d,c,t} \quad (2.1)$$

This equation is similar to a traditional differences-in-differences design, but instead of looking at a single treatment indicator, we now have a number of treatment leads and lags corresponding to the time until (or since) treatment. The β_i coefficients are the leads of policy adoption. We would expect these to be zero since the policy should not affect outcomes in previous years. A possible exception to this might be β_1 since there might be differential attrition in anticipation of the policy change depending on when it was announced. The coefficients contained within β_j (lags) are the coefficients of interest.

- y is an outcome for teacher i in school s district d county c and school year t
- δ_t captures time fixed effects
- δ_s captures school-level fixed effects that are time-invariant (we shouldn't need this)
- β_j are coefficients of interest
- β_i are expected to be zero since the policy shouldn't affect the pre-treatment period

The event study can be used with a variety of control groups. First, we can use eventual adopters since these seem like the most natural counterfactual group to compare the treated teachers/schools to. However, this can present a couple of problems. To begin with, this greatly shrinks our sample size and reduces our statistical power. Additionally, recent research into staggered treatment designs has revealed that TWFEs under staggered treatment timing can be biased because of how the average treatment effect is generated (Baker, Larcker, and Wang, 2022). Therefore, as a robustness check, I also use schools that are similar to treated schools – schools given the NCES classification of "rural" or "town" since nearly all four-day schools share this classification – but never receive treatment during the years spanned by the panel dataset.

2.5 Results

Figure 2.4 reflects the attrition rate for all full-time teachers in Oklahoma. Since teacher attrition is at least partly a function of opportunity wages, quit rates declined during the 2008 financial crisis, but have been steadily climbing upwards since. However, attrition is also largely a function of experience with attrition rates declining as one acquires teaching experience, and in turn, tenure, in addition to climbing the salary schedule and becoming vested for retirement. Figure 2.5 shows how these experience-specific attrition rates have evolved over time. In addition, teacher turnover can vary based on school locale. Whereas attrition rates were quite similar across all schools 15-20 years ago, Figure 2.6 demonstrates that this is no longer the case, with schools located in cities having unconditional quit rates 30 percent higher than those located in non-urban areas.

The results of our event study design are contained in Table 2.1. Our dependent variable is a binary indicator that is equal to zero if a teacher returns to teach the following year and one if the teacher is not seen in the data in future years. This setup yields a linear probability model of attrition for each teacher in a given year or, when aggregated, annual attrition at the school or district level. Because the baseline attrition rate for teachers is both time- and district-dependent, as evidenced above, I include year and district fixed effects in all specifications. Attrition is largely a function of teacher experience, so experience bins are included in all regressions. In addition, these regressions also include a full set of teacher-, school-, and county-level controls. Column 1 is our regression model that includes all of the full-time teachers who have yet to reach retirement age in the years spanned by our panel. Column 2 includes all teachers that teach at non-urban schools since these are very different than the more rural schools where the four-day school weeks are concentrated.[11] Column 3 uses only "eventually treated" schools as a counterfactual, which has the benefit of accounting for the idiosyncrasies of adopting schools but also comes at the cost of greatly reducing our sample size. Across all three of these models, we are not able to detect any statistically significant effect on attrition until year 2, where we see a meaningful reduction in quit rates of 1.6 percentage points or 25 percent. 25 percent is certainly not a small reduction in attrition, and one might expect this effect to be partially attenuated by the number of teachers for whom the policy will likely not affect them, e.g., someone with 15 years of service (roughly the average in four-day districts).

[11] There are zero "city" schools in Oklahoma that have four-day school weeks

These delayed effects observed in teacher attrition post-treatment are further bolstered by the reduction in the number of emergency certificates seen in Table 2.4 and the number of emergency certificated subject areas in Table 2.5. The outcome variable in these tables and figures is constructed by dividing the number of emergency certificates (or subject areas) required by each district by that district's number of new hires. This metric like a reasonable way of getting at the effect the policy is having on a school's ability to improve their hiring. The reduction in usage of emergency-certificated teachers is not seen until year three (Figure 2.10 and 2.11); however, since a great majority of emergency certificates are renewed annually (up to 3 years) until they expire, we could reasonably expect to see the usage of these measures slowly decline as marginal teachers transition to permanent teaching credentials or need is reduced with the addition of new hires. By the third year, we see a reduction in our outcome measure of 3.1 percent, which implies that for roughly every 30 new hires, an emergency credential is avoided.

Lastly, I also find evidence regarding the selection – recruitment and retention – effects of this policy. First, using prior teaching experience as a proxy for teaching quality, Figure 2.13 shows how schools' ability to hire experienced teachers improves after four-day school week adoption. Three years into the policy, these schools are able to increase the proportion of new hires with prior experience by 2.7 percentage points versus similarly located schools.[12] Next, now looking only at teachers with zero previous teaching experience in Figure 2.12, we can see that any school-wide retention effects are

[12] A surprisingly high percentage (65 %) of new hires in rural schools have previous teaching experience.

primarily being driven by these brand-new teachers quitting at much lower rates during their first year (6.1 percentage points). Given the baseline first-year attrition rate of 13.7 percent, this implies a reduction in first-year attrition of nearly 45 percent.

2.6 Conclusion

This paper attempts to empirically estimate the effect of four-day school weeks on schools' ability to hire and retain teachers. The evidence from 15 years of Oklahoma employee data suggests that these four-day schedule changes induce more interest in teaching at treated schools, both among new teachers and those with previous teaching experience at other schools. I observe no negative effects on the existing cohorts that are working at these schools prior to the schedule change.

Furthermore, I find evidence suggestive of an improvement in teacher quality. Of course, as mentioned before, a school's ability to hire teachers with previous teaching experience is likely associated with an increase in teacher quality given the returns to experience (Harris and Sass, 2011; Rice, Croninger, and Roellke, 2002). However, in addition to this potential benefit, we also see decreased need at treated schools for emergency credentials for new teachers.

These findings are particularly interesting given another late but quick adopter of four-day school weeks: Missouri. Missouri is a curious four-day week case study because of both the rate and scale of adoption – the number of adopting districts has quadrupled in the last 4 years, with more than 25 percent of districts now only operating four days per week – but also because districts are unabashedly using it as a way of leveling the playing field with regard to teacher recruitment and retention. In many ways, Missouri mirrors Oklahoma, but what is of particular interest from a policy analysis perspective is the rise in test scores in these schools. Given the findings on the student achievement effects, one might expect to see test scores fall in Missouri, but Paul N. Thompson and Ward (2022) finds increases of nearly .05 standard deviations in both reading and math achievement after the schedule switch.[13] While the majority of schools have seen a reduction in test scores associated with the schedule change, it is notable that test scores have improved in the state that has leaned into four-day weeks as a teacher recruitment tool.[14] One thing is clear: in spite of the recent evidence that four-day school weeks, on average, hurt student achievement, there doesn't seem to be anything slowing down increased adoption.

[13] It's not clear if this is important, but it bears mentioning that most four-day schools in Missouri have chosen to drop Mondays instead of Fridays.

[14] Thompson also argues that this is partly due to increased instructional time, but that is relative to other adopting states which don't require as many instructional hours as Missouri. However, Colorado has more instructional time than any other 4DSW state, and even it has seen neutral to negative effects of adoption.

References

Baker, Andrew C, David F Larcker, and Charles CY Wang (2022). "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2, pp. 370–395.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor (2011). "Teacher mobility, school segregation, and pay-based policies to level the playing field". In: *Education Finance and Policy* 6.3, pp. 399–438.

Cross, Freddie (2017). *Teacher Shortage Areas Nationwide Listing 1990–1991 through 2017–2018 June 2017.* Tech. rep.

Garcia, Emma and Elaine Weiss (2019). "The Teacher Shortage Is Real, Large and Growing, and Worse than We Thought. The First Report in" The Perfect Storm in the Teacher Labor Market" Series." In: *Economic Policy Institute.*

Goldhaber, Dan and Michael Hansen (2010). "Using performance on the job to inform teacher tenure decisions". In: *American Economic Review* 100.2, pp. 250–55.

Hamermesh, Daniel S and Jeff Biddle (2022). *Days of Work over a Half Century: The Rise of the Four-Day Week.* Tech. rep. National Bureau of Economic Research.

Hanushek, Eric A and Ludger Woessmann (2020). “The economic impacts of learning losses”. In.

Harris, Douglas N. and Tim R. Sass (2011). “Teacher training, teacher quality and student achievement”. In: *Journal of Public Economics* 95, pp. 798–812.

Hendricks, Matthew D (2014). “Does it pay to pay teachers more? Evidence from Texas”. In: *Journal of Public Economics* 109, pp. 50–63.

Hunt, R (1936). “A fusion high school program”. In: *American School Board Journal* 93.5, p. 46.

Ingersoll, Richard (2003). “Is there really a teacher shortage?” In.

Johnson, David P (1977). “Four-Day Work Week Saves Energy.” In: *Community and Junior College Journal* 48.3, pp. 31–3.

Kordosky, D. (2011). *The Four Day School Week: Less IS More!* Dog Ear Publishing, LLC. ISBN: 9781457504228. URL: https://books.google.com/books?id=S7kOywAACAAJ.

Learning Policy Institute (May 2022). *Uncertified teachers and teacher vacancies by state*. Tech. rep. URL: https://learningpolicyinstitute.org/uncertified-teachers-and-teacher-vacancies-state.

Miller, Raegen T., Richard J. Murnane, and John B. Willett (2008). “Do Teacher Absences Impact Student Achievement? Longitudinal Evidence From One Urban School District”. In: *Educational Evaluation and Policy Analysis* 30.2, pp. 181–200. DOI: 10.3102/0162373708318019.

Morton, Emily (2021). “Effects of four-day school weeks on school finance and achievement: Evidence from Oklahoma”. In: *Educational Researcher* 50.1, pp. 30–40.

Nguyen, Tuan, Chanh Lam, and Paul Bruno (2022). “Is there a national teacher shortage? A systematic examination of reports of teacher shortages in the United States.(EdWorkingPaper: 22-631)”. In.

Pedersen, James (2012). “The History of School and Summer Vacation.” In: *Journal of Inquiry and Action in Education* 5.1, pp. 54–62.

Rice, Jennifer King, Robert G Croninger, and Christopher F Roellke (2002). “The effect of block scheduling high school mathematics courses on student achievement and teachers’ use of time: implications for educational productivity”. In: *Economics of Education Review* 21.6, pp. 599–607.

Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff (2013). “How teacher turnover harms student achievement”. In: *American educational research journal* 50.1, pp. 4–36.

Rosen, Sherwin (1986). “The theory of equalizing differences”. In: *Handbook of labor economics* 1, pp. 641–692.

Rowland, Julie (2014). “Number of Instructional Days/Hours in the School Year.” In: *Education Commission of the States.*

Salop, Steven C (1979). “A model of the natural rate of unemployment”. In: *The American Economic Review* 69.1, pp. 117–125.

Schaeffer, Katherine (Sept. 2022). “A dwindling number of new U.S. college graduates have a degree in education a dwindling number of new U.S. college graduates have a degree in Education”. In: *Pew Research Center*.

Schlicht, Ekkehart (1978). “Labour turnover, wage structure, and natural unemployment”. In: *Zeitschrift für die Gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, pp. 337–346.

Stiglitz, Joseph E (1976). “The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs”. In: *Oxford economic papers* 28.2, pp. 185–207.

Sutcher, Leib, Linda Darling-Hammond, and Desiree Carver-Thomas (2019). “Understanding teacher shortages: An analysis of teacher supply and demand in the United States.” In: *Education Policy Analysis Archives* 27.35.

Tharp, Timothy W (2014). *A comparison of student achievement in rural schools with four and five day weeks*. University of Montana.

Thompson, Paul N (2021). “Is four less than five? Effects of four-day school weeks on student achievement in Oregon”. In: *Journal of public economics* 193, p. 104308.

Thompson, Paul N. and Jason Ward (2022). “Only a matter of time? The role of time in school on four-day school week achievement impacts”. In: *Economics of Education Review* 86, p. 102198. ISSN: 0272-7757. DOI: https://doi.org/10.1016/j.econedurev.2021.102198. URL: https://www.sciencedirect.com/science/article/pii/S0272775721001138.

Turner, Jon S, Kim K Finch, and Ximena Uribe-Zarain (2018). “Staff perspectives of the four-day school week: A new analysis of compressed school schedules”. In: *Journal of Education and Training Studies*.

Walsh, Kate (Dec. 2016). “The national teacher shortage is a myth. here’s what’s really happening.” In: *The Washington Post*.

2.8 Figures and tables

Figure 2.1: States that have used four-day weeks (as of 2019)

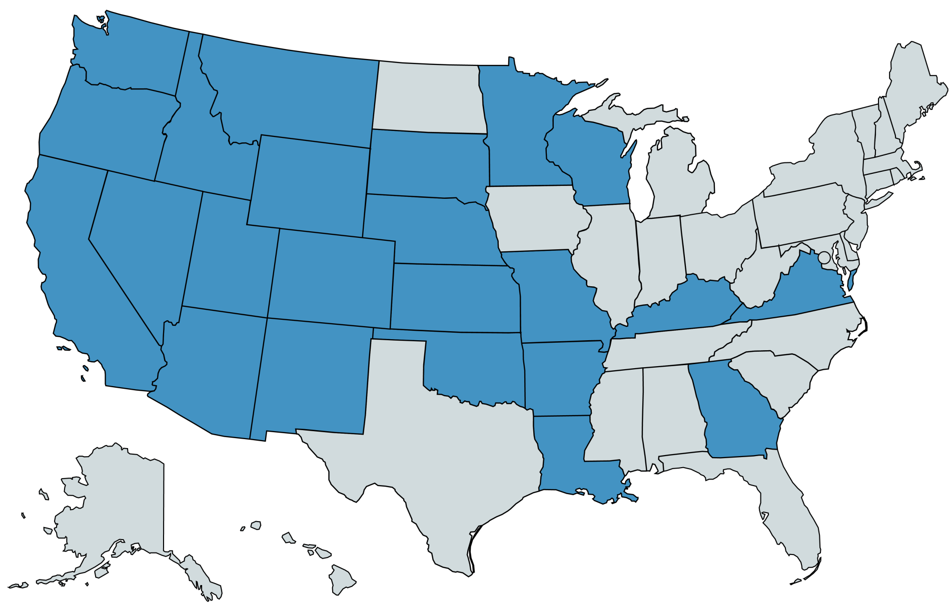
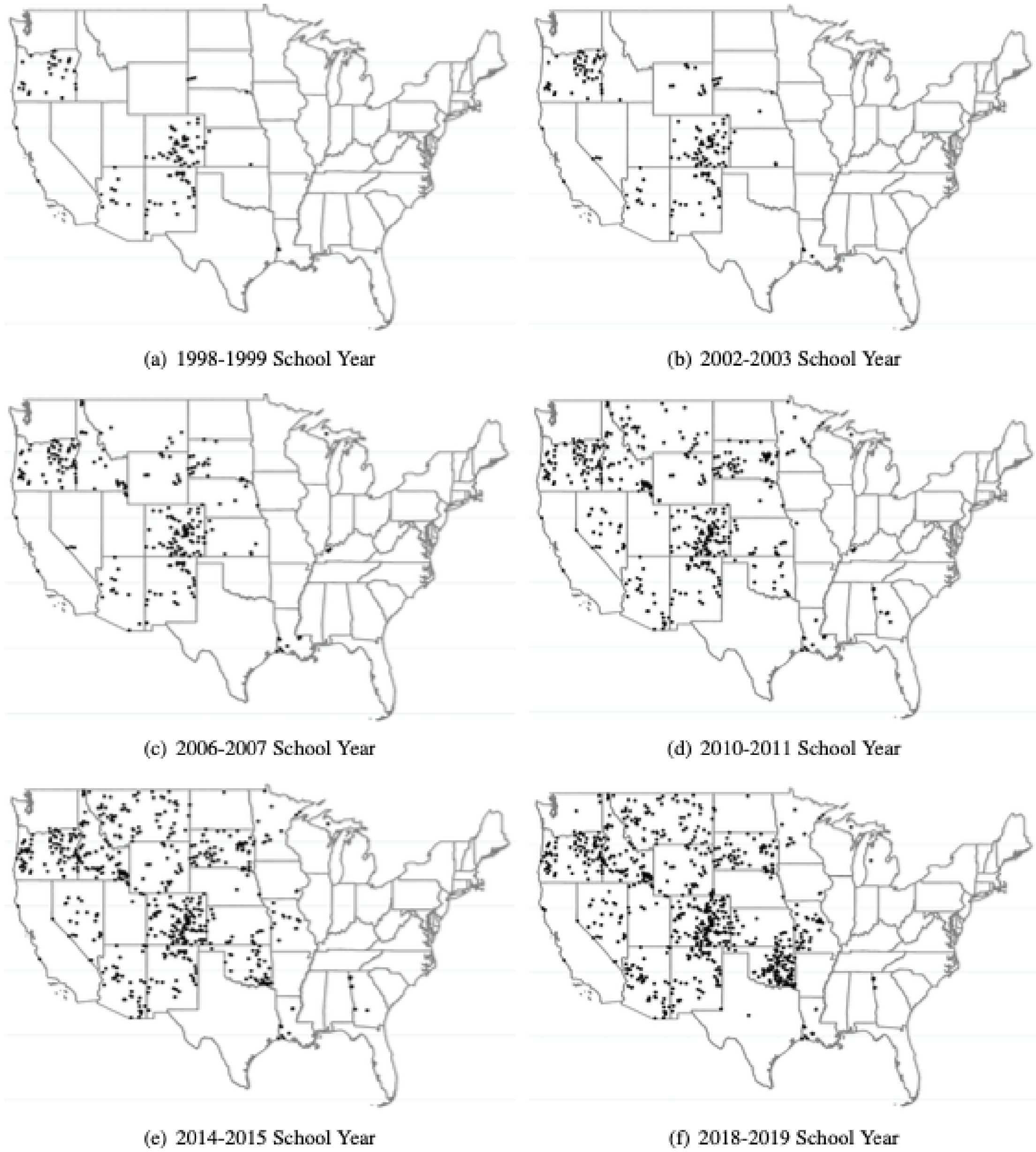


Figure 2.2: Map of four-day school week adoption by year - National



(a) 1998-1999 School Year

(b) 2002-2003 School Year

(c) 2006-2007 School Year

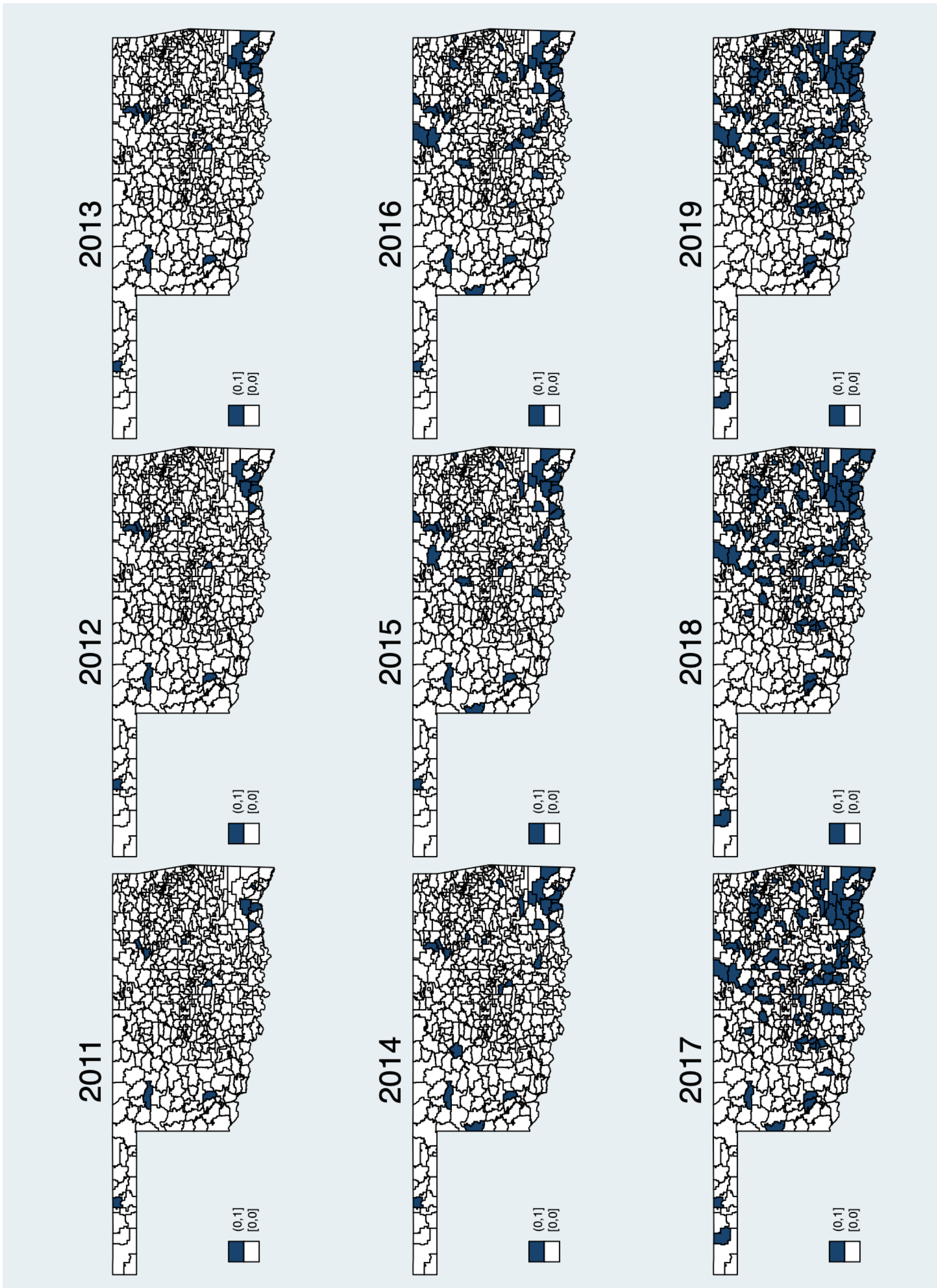
(d) 2010-2011 School Year

(e) 2014-2015 School Year

(f) 2018-2019 School Year

Notes: This panel indicates the FDSW adoption status across the United States from 1998 to 2019. Dots are used to indicate an adopting district.

Figure 2.3: Map of four-day school week adoption by year - Oklahoma



This panel indicates the FDSW adoption status across all 509 school districts from 2011 to 2019. Districts are shaded blue to indicate adoption of the policy.

Figure 2.4: Attrition rate for full-time teachers in Oklahoma by year

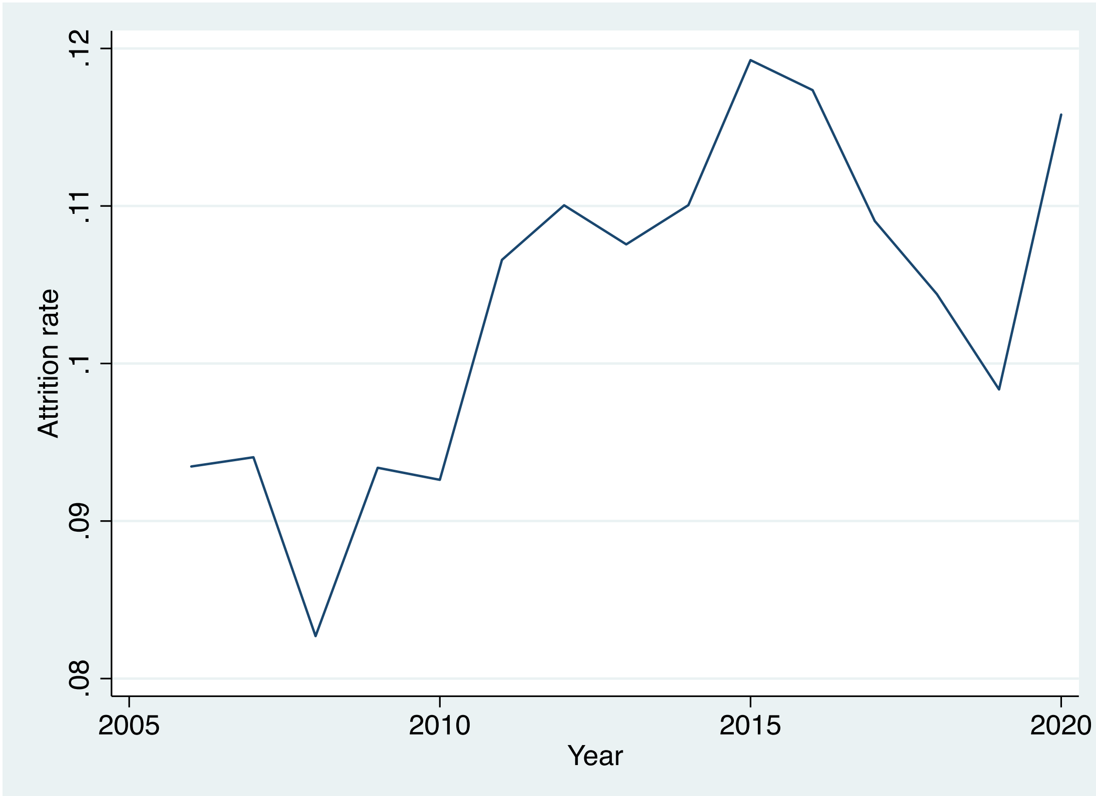


Figure 2.5: Attrition rate for full-time teachers in Oklahoma by year and experience bin

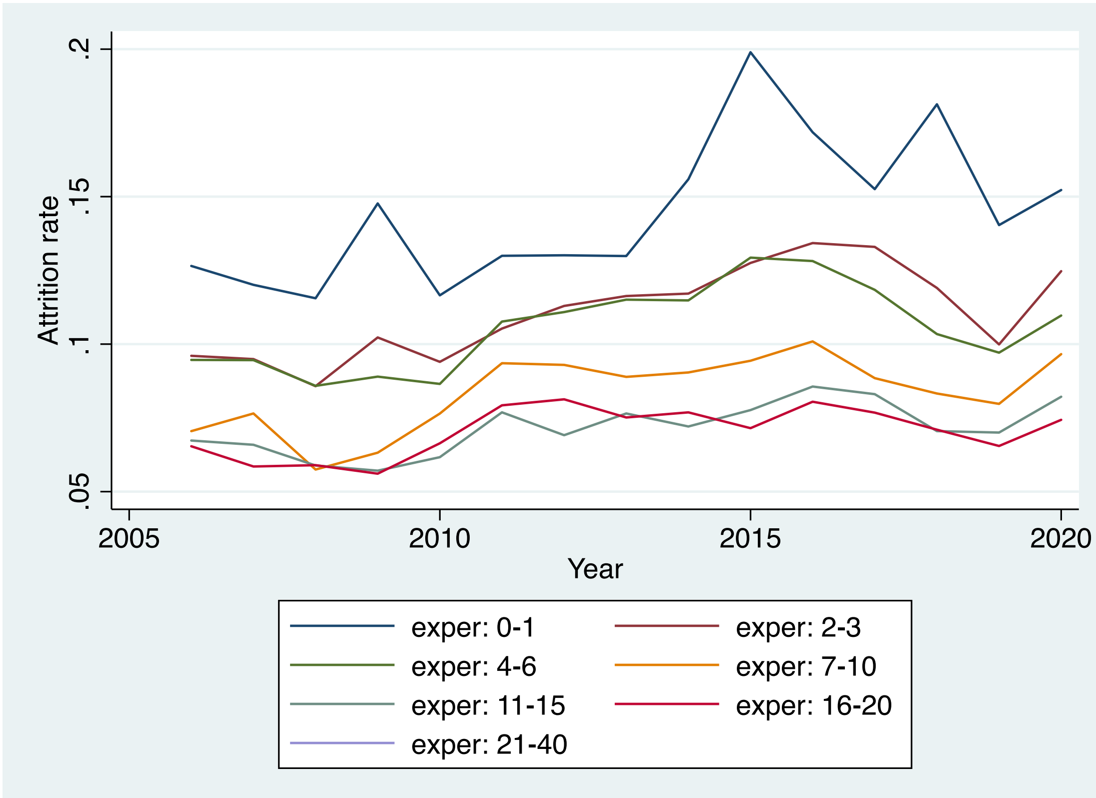
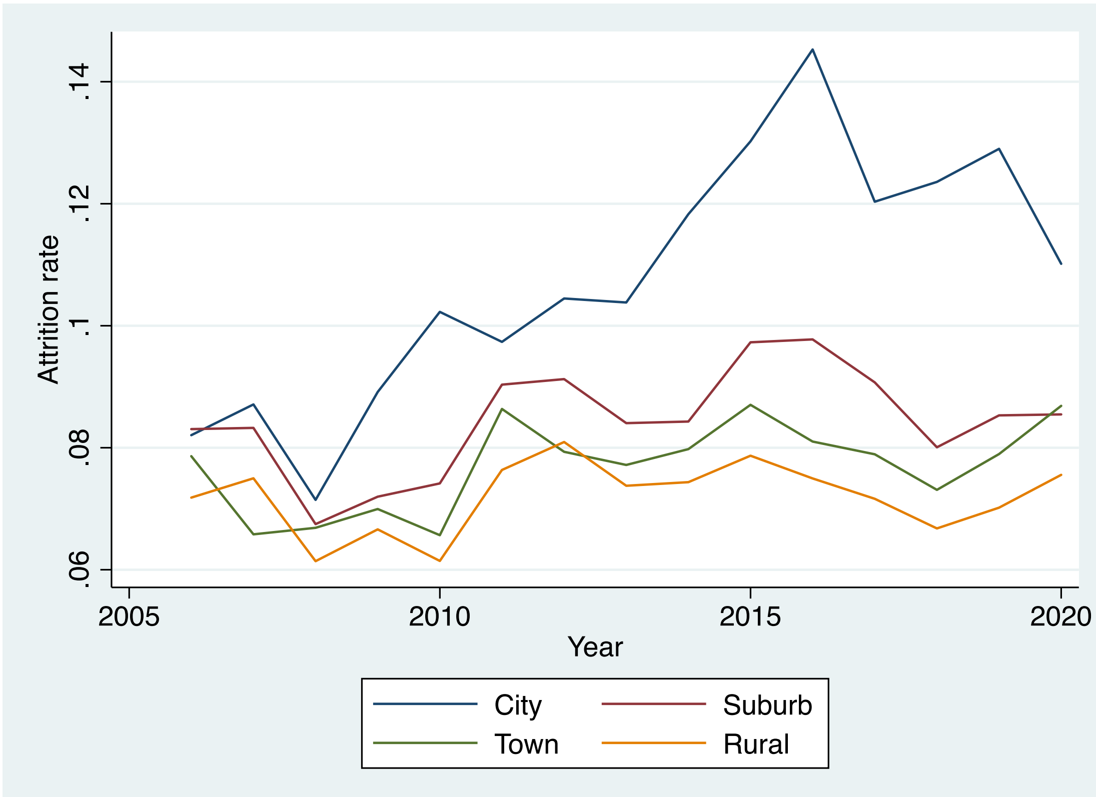


Figure 2.6: Attrition rate for full-time teachers in Oklahoma by year and locale



Notes: For a description of school location classification methodology, see: nces.ed.gov/programs/edge/docs/locale$_{classifications.pdf}$

Figure 2.7: Emergency teacher certificates issued by year

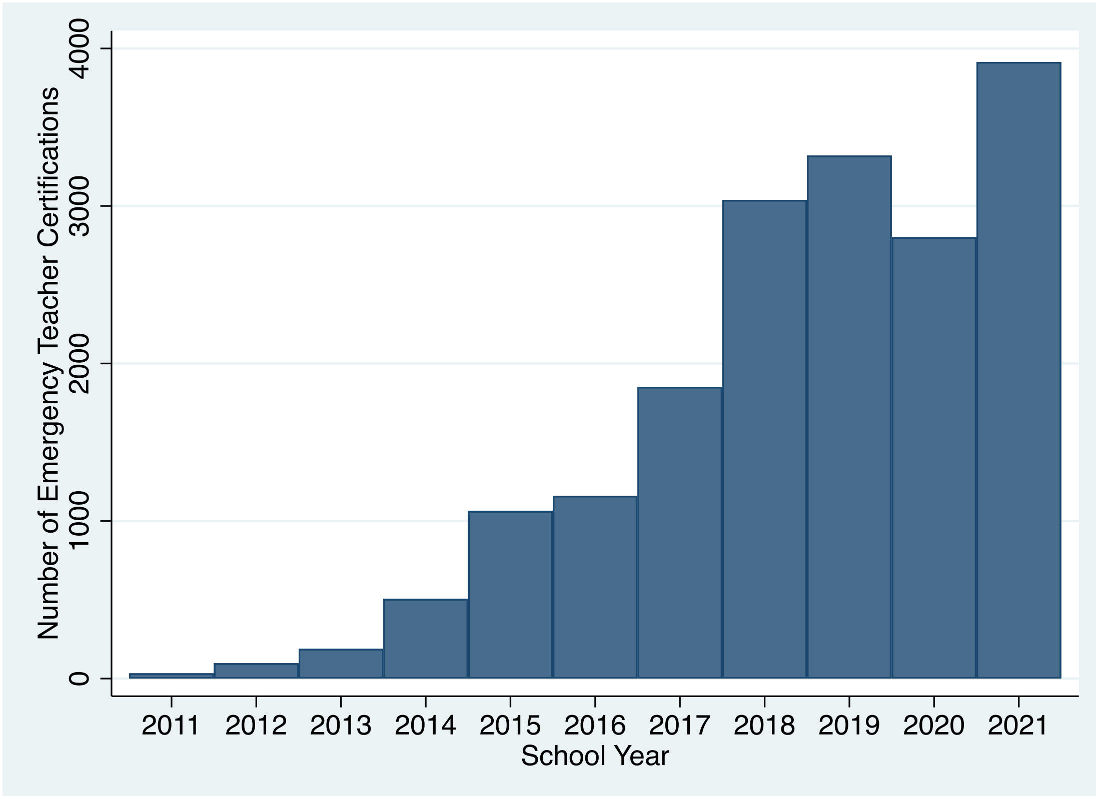
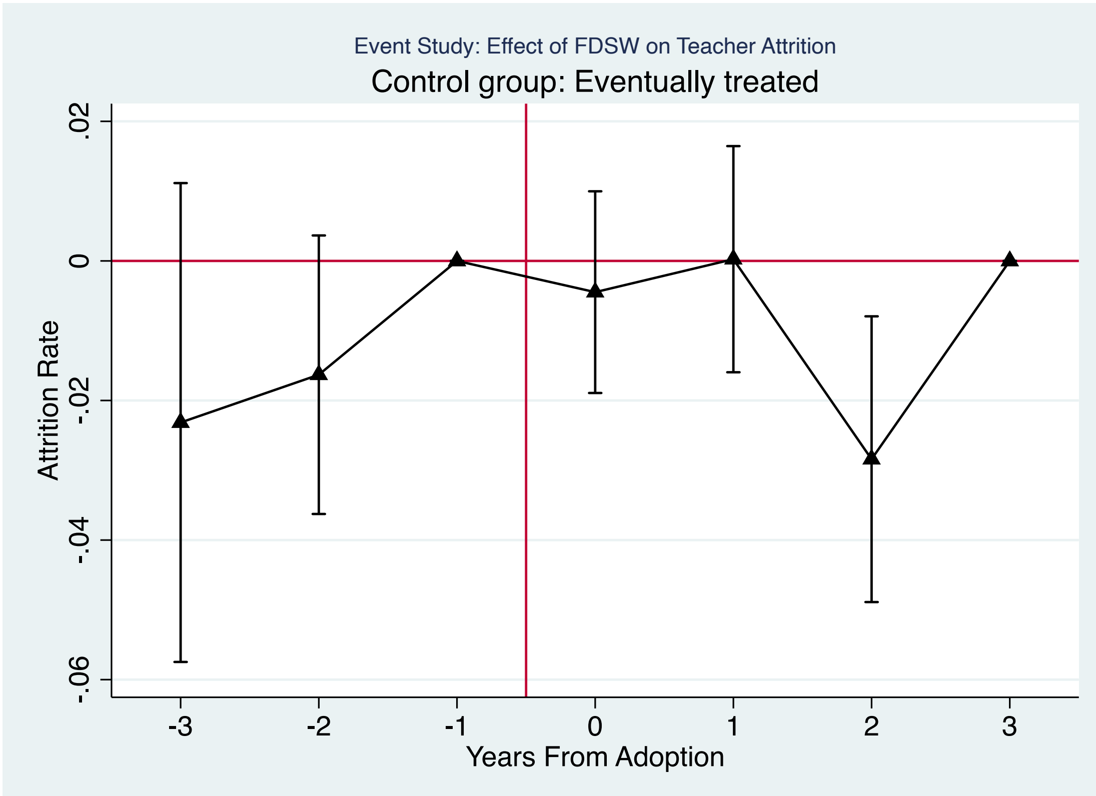
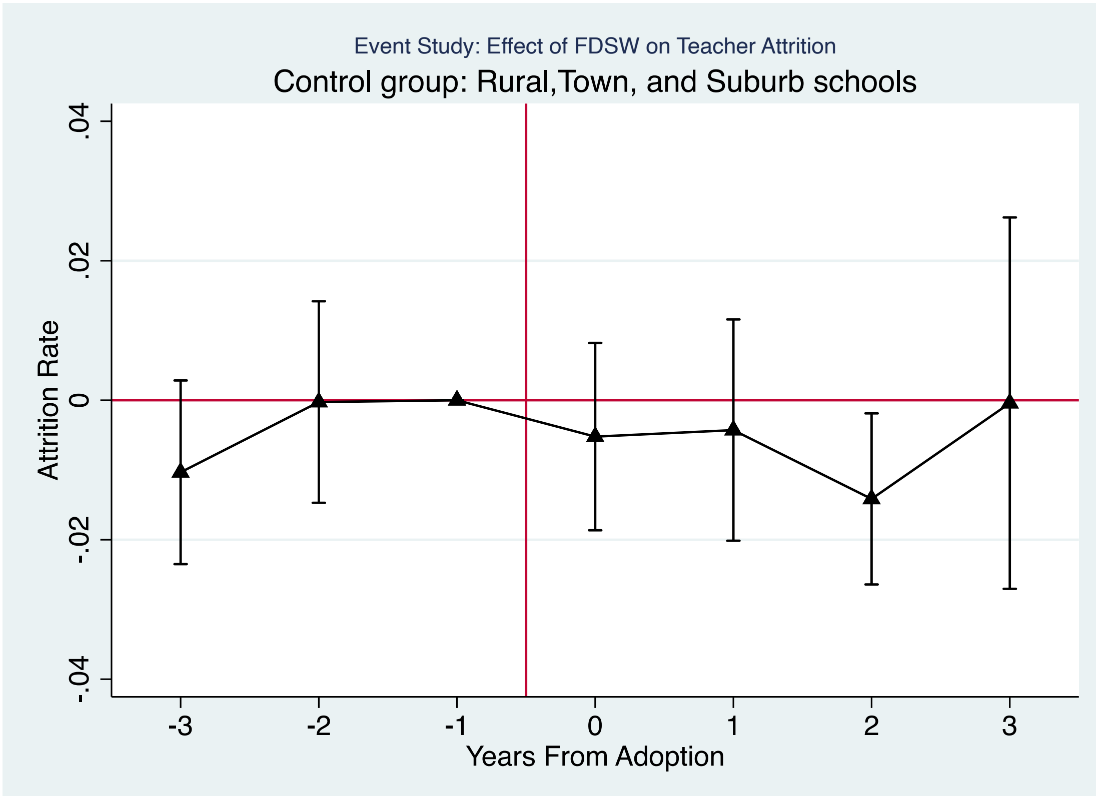


Figure 2.8: Mean difference in residual attrition rate between treatment and control schools



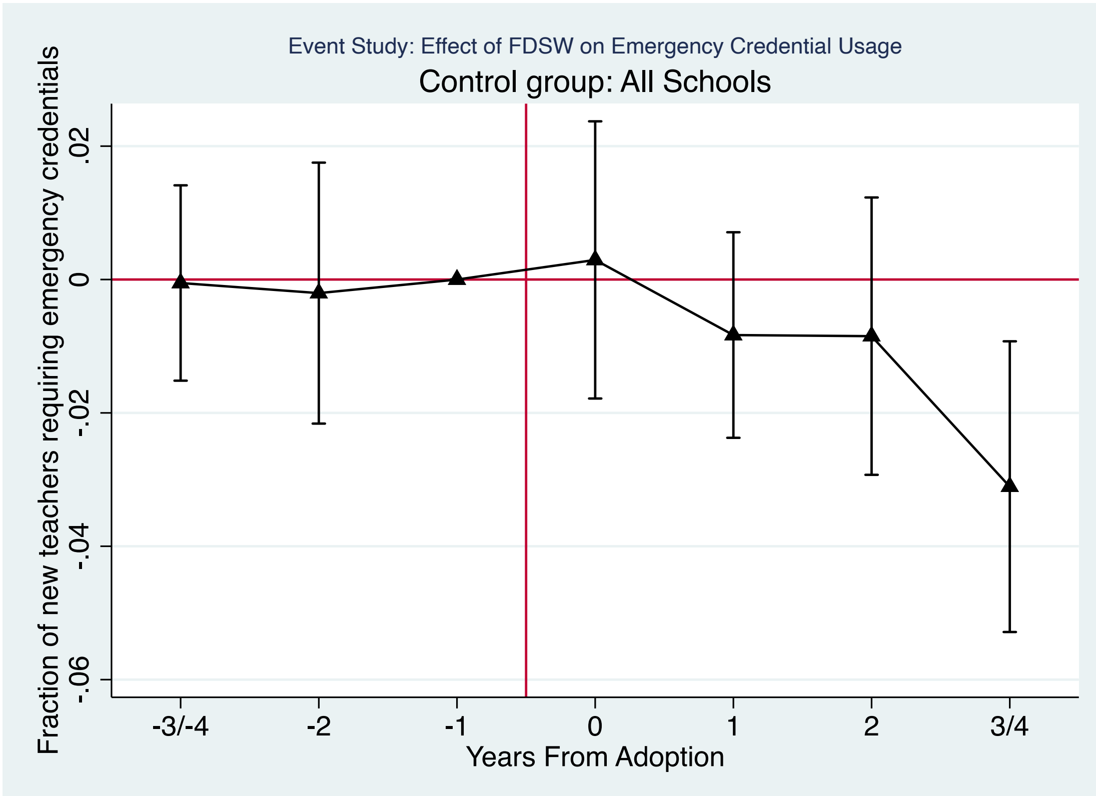
Notes: Figure 2.8 plots the event time indicators relative to time of treatment from Model 4 in Table 2.2

Figure 2.9: Mean difference in residual attrition rate between treatment and control schools



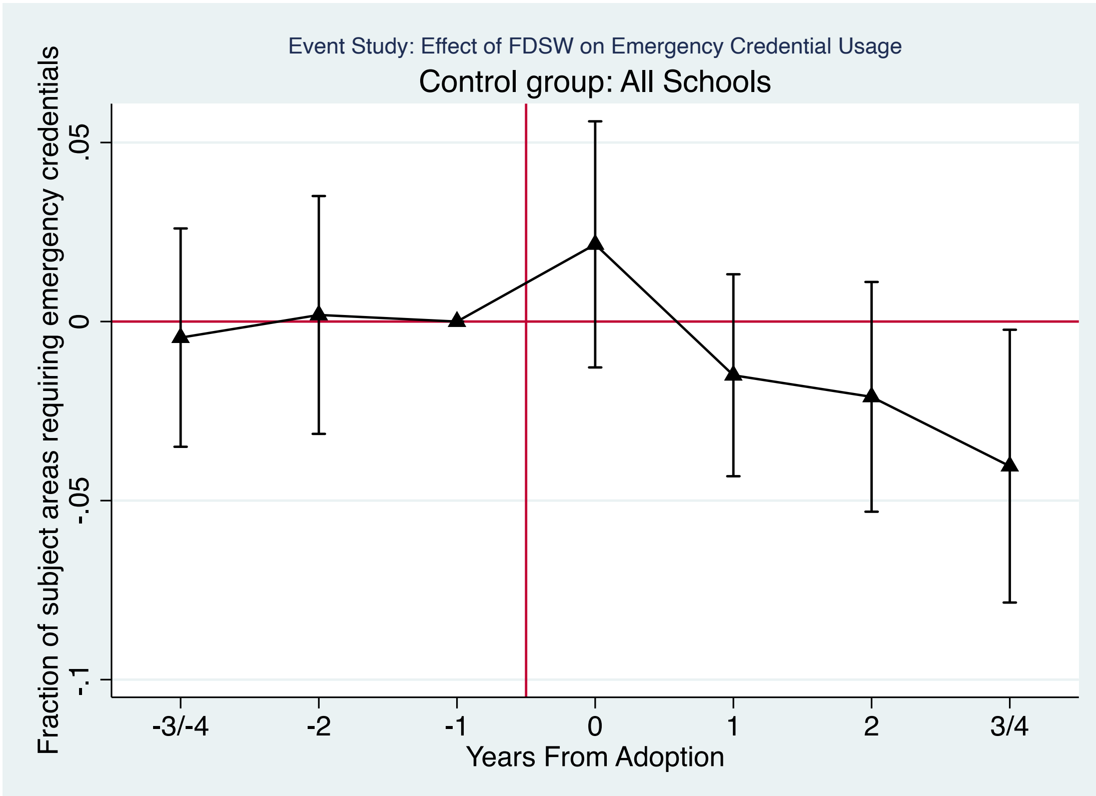
Notes: Figure 2.9 plots the event time indicators relative to time of treatment from Model 4 in Table 2.3

Figure 2.10: Mean difference in residual emer. cert. rate between treatment and control schools



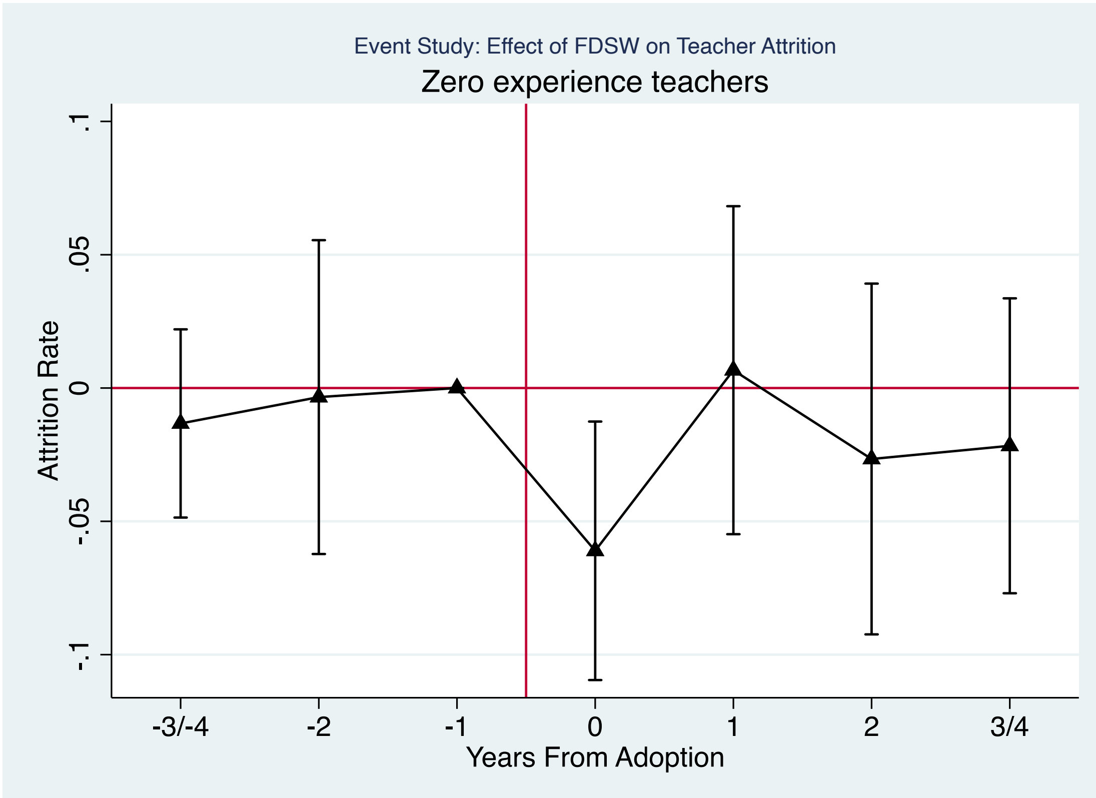
Notes: Figure 2.10 plots the event time indicators relative to time of treatment from Model 4 in Table 2.4

Figure 2.11: Mean difference in residual emer. cert. rate between treatment and control schools



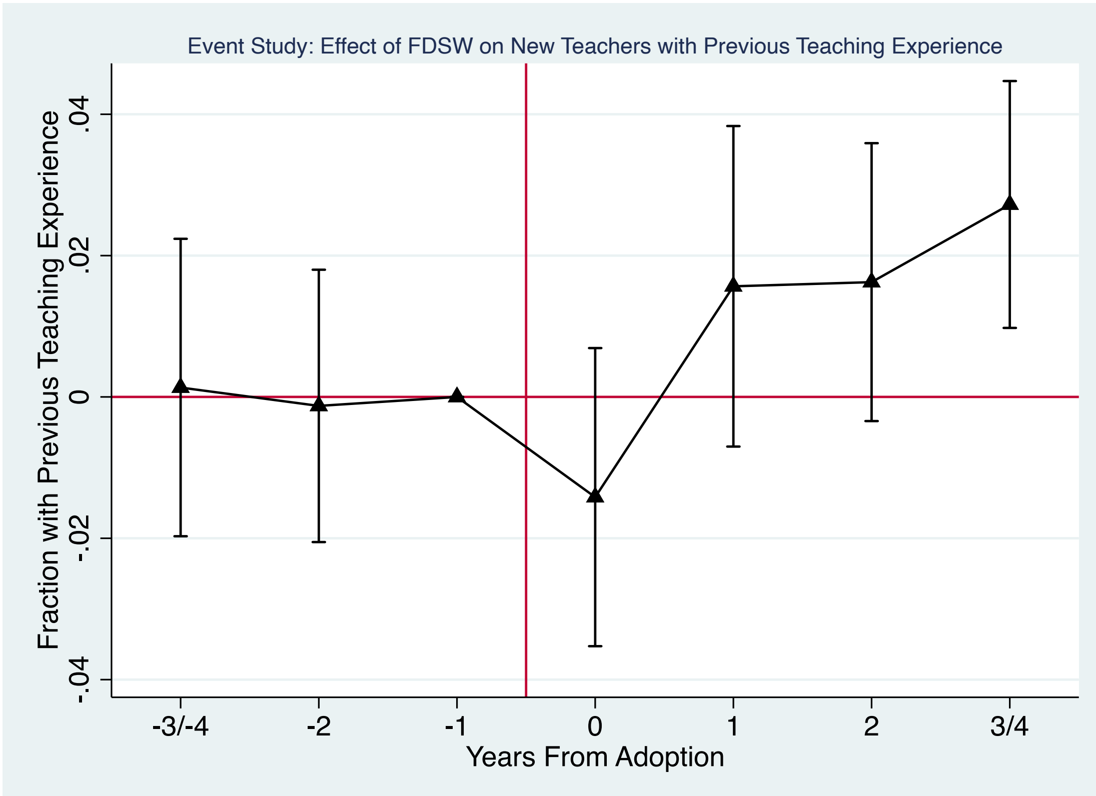
Notes: Figure 2.11 plots the event time indicators relative to time of treatment from Model 4 in Table 2.5

Figure 2.12: Mean difference in residual attrition rate between treatment and control schools



Notes: Figure 2.12 plots the event time indicators relative to time of treatment from Model 4 in Table 2.6

Figure 2.13: Mean difference in residual attrition rate between treatment and control schools



Notes: Figure 2.13 plots the event time indicators relative to time of treatment from Model 4 in Table 2.7

Table 2.1: Event Study: Effect of four-day school week on teacher attrition w/ different controls

	(1)	(2)	(3)
3-4 years before	0.00144	0.00188	-0.000250
	(0.00430)	(0.00429)	(0.00474)
2 years before	-0.00571	-0.00397	-0.00782
	(0.00522)	(0.00511)	(0.00576)
Year of adoption	-0.00817	-0.00239	-0.00457
	(0.00576)	(0.00543)	(0.00669)
1 year after	0.00399	0.00619	0.0135*
	(0.00610)	(0.00613)	(0.00703)
2 years after	-0.00819	-0.00531	-0.0156*
	(0.00646)	(0.00623)	(0.00792)
3-4 years after	0.00208	0.00519	-0.00349
	(0.00474)	(0.00432)	(0.00617)
Year FE	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Observations	463737	358808	37642
Mean of Dep. Variable	0.0762	0.0688	0.0635

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Testing notes

Table 2.2: Event Study: Effect of four-day school week on teacher attrition using eventually treated schools

	(1)	(2)	(3)	(4)
3-4 years before	0.00216	-0.00285	-0.00161	0.000434
	(0.00424)	(0.00484)	(0.00471)	(0.00466)
2 years before	-0.00351	-0.00885	-0.00824	-0.00573
	(0.00570)	(0.00652)	(0.00637)	(0.00632)
Year of adoption	0.00305	-0.00291	-0.00232	-0.00129
	(0.00555)	(0.00637)	(0.00631)	(0.00595)
1 year after	0.00602	0.00327	0.00367	0.00450
	(0.00558)	(0.00670)	(0.00643)	(0.00645)
2 years after	-0.00228	-0.0116	-0.00875	-0.00906
	(0.00615)	(0.00809)	(0.00784)	(0.00751)
3-4 years after	0.00171	-0.0124*	-0.00744	-0.00592
	(0.00559)	(0.00707)	(0.00655)	(0.00613)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	56243	56243	56243	55854
Mean of Dep. Variable	0.0959	0.0959	0.0959	0.0952

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Dependent variable: Left teaching. Coefficients represent the effect of time relative to policy adoption. Models 3 and 4 include experience bins. Model 4 includes full set of controls. Standard errors are clustered at the district level.

Table 2.3: Event Study: Effect of four-day school week on teacher attrition using non-urban schools as controls

	(1)	(2)	(3)	(4)
3-4 years before	-0.00714	-0.00844	-0.00707	-0.00636
	(0.00616)	(0.00639)	(0.00621)	(0.00617)
2 years before	0.00144	0.000443	0.00157	0.00263
	(0.00706)	(0.00722)	(0.00662)	(0.00655)
Year of adoption	-0.00344	-0.00784	-0.00645	-0.00543
	(0.00653)	(0.00692)	(0.00625)	(0.00621)
1 year after	-0.000644	-0.00359	-0.00224	-0.00191
	(0.00674)	(0.00697)	(0.00704)	(0.00694)
2 years after	-0.0137**	-0.0129**	-0.0104*	-0.0103*
	(0.00543)	(0.00555)	(0.00554)	(0.00552)
3 years after	0.00458	0.00426	0.00937	0.00973
	(0.00869)	(0.00884)	(0.00752)	(0.00749)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	332153	332153	332151	332105
Mean of Dep. Variable	0.0769	0.0769	0.0769	0.0769

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Event Study: Effect of four-day school week on emergency certificate usage

	(1)	(2)	(3)	(4)
3-4 years before	-0.0243***	0.0120**	0.000368	-0.000520
	(0.00804)	(0.00537)	(0.00856)	(0.00745)
2 years before	-0.00962	0.0193*	0.0000606	-0.00204
	(0.0126)	(0.0113)	(0.0113)	(0.00995)
Year of adoption	0.0297	0.0427**	0.00648	0.00294
	(0.0185)	(0.0168)	(0.0116)	(0.0106)
1 year after	0.00229	0.00986	-0.00181	-0.00832
	(0.0113)	(0.0101)	(0.00941)	(0.00784)
2 years after	0.0112	-0.00162	-0.0106	-0.00849
	(0.0107)	(0.0112)	(0.00949)	(0.0106)
3-4 years after	0.0118	-0.00468	-0.0120	-0.0311***
	(0.0103)	(0.0136)	(0.00765)	(0.0111)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	2226	2226	2170	1345
Mean of Dep. Variable	0.0668	0.0668	0.0613	0.0549

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Outcome of interest is defined as the "number of emergency certificates issued to a district in a given year divided by the number of new teachers hired that year"

Table 2.5: Event Study: Effect of four-day school week on emergency certificate usage

	(1)	(2)	(3)	(4)
3-4 years before	-0.0253**	0.0143	-0.00380	-0.00449
	(0.0115)	(0.00972)	(0.0151)	(0.0155)
2 years before	-0.000177	0.0328*	-0.00177	0.00184
	(0.0185)	(0.0178)	(0.0226)	(0.0169)
Year of adoption	0.0565**	0.0729***	0.0242	0.0215
	(0.0235)	(0.0217)	(0.0182)	(0.0175)
1 year after	0.0134	0.0209	0.000690	-0.0150
	(0.0192)	(0.0185)	(0.0169)	(0.0143)
2 years after	0.0101	-0.00535	-0.0201	-0.0210
	(0.0132)	(0.0143)	(0.0146)	(0.0163)
3-4 years after	0.0137	-0.00354	-0.0143	-0.0404**
	(0.0124)	(0.0161)	(0.0115)	(0.0194)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	2226	2226	2170	1345
Mean of Dep. Variable	0.0836	0.0836	0.0774	0.0706

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Outcome of interest is defined as the "number of emergency certificate subject areas issued to a district in a given year divided by the number of new teachers hired that year".

Table 2.6: Event Study: Effect of four-day school week on teacher attrition with zero prior experience

	(1)	(2)	(3)	(4)
3-4 years before	0.00502	0.0124	-0.0166	-0.0133
	(0.0184)	(0.0186)	(0.0191)	(0.0180)
2 years before	0.0150	0.0214	-0.00576	-0.00340
	(0.0267)	(0.0270)	(0.0304)	(0.0300)
Year of adoption	-0.0198	-0.0289	-0.0610**	-0.0611**
	(0.0256)	(0.0254)	(0.0256)	(0.0247)
1 year after	0.0593**	0.0578**	0.0188	0.00669
	(0.0284)	(0.0292)	(0.0316)	(0.0313)
2 years after	0.00396	-0.00663	-0.0301	-0.0266
	(0.0288)	(0.0281)	(0.0334)	(0.0335)
3-4 years after	0.00918	0.0105	-0.0282	-0.0217
	(0.0274)	(0.0273)	(0.0308)	(0.0282)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	29591	29591	29584	29575
Mean of Dep. Variable	0.137	0.137	0.137	0.137

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Event Study: Effect of four-day school week on school's ability to hire experienced teachers

	(1)	(2)	(3)	(4)
3-4 years before	0.0342	0.0206	0.00208	0.00133
	(0.0210)	(0.0216)	(0.0106)	(0.0107)
2 years before	0.0354*	0.0307	-0.000568	-0.00127
	(0.0209)	(0.0219)	(0.00992)	(0.00981)
Year of adoption	0.0218	0.0162	-0.0128	-0.0142
	(0.0252)	(0.0258)	(0.0108)	(0.0107)
1 year after	0.0716***	0.0570**	0.0159	0.0157
	(0.0243)	(0.0249)	(0.0115)	(0.0115)
2 years after	0.0626***	0.0594***	0.0169*	0.0162
	(0.0212)	(0.0220)	(0.0100)	(0.0100)
3-4 years after	0.0760***	0.0689***	0.0289***	0.0272***
	(0.0188)	(0.0191)	(0.00887)	(0.00889)
Year FE	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Observations	52146	52146	52143	52121
Mean of Dep. Variable	0.652	0.652	0.652	0.652

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

The "Hazards" of Four-Day School Weeks: Tracking Teacher Cohorts in Oklahoma

3.1 Introduction

The decision to enter or remain in the labor market for teachers before each school year is a difficult one faced by millions of educators every year. It is an unusual profession because of the extraordinarily high attrition rate in the early years – well more than half (60 percent) of teachers that started teaching in Oklahoma in 2012 were still teaching in the state five years later (Oklahoma State Dept. of Education, 2021).

As discussed in Barber (2022), the shortage of teachers is less of a national problem and more of a regional one, and Oklahoma has certainly faced its own struggles in maintaining an adequate supply of qualified teachers. Much of this is a function of the salaries that teachers receive in Oklahoma schools – the state resides at the bottom of the rankings across most salary metrics.[1]

The undersupply of teachers is a multifaceted problem, but much of this is a function of the salaries available to those wanting to pursue (or continue) a career in education. Teaching salaries actually **declined** in real terms in many states, including Oklahoma, in the ten-year period after the 2008 financial crisis (Oklahoma State Dept. of Education, 2021; NEA Research, 2022). Simultaneously, teachers are paid 22 percent less than they would be if they switched to jobs outside of teaching (Hanushek, Piopiunik, and Wiederhold, 2019). Naturally, states are limited in their ability to increase pay for teachers, constrained by property tax revenues, state budgets, and federal support.[2]. It is this against this austerity backdrop that states like Oklahoma have begun looking at non-monetary compensation options, like four-day school weeks, to retain existing faculty and entice both new and previously tenured teachers.[3]

Furthermore, the structure of teacher compensation, specifically the balance between salary and benefits, in Oklahoma is likely discouraging to the marginal teacher. Both Biasi (2019) and Johnston (2021) find that teachers are much more responsive to

[1]Prior a recent pay raise, Oklahoma was ranked 49th in average teacher salary, but has now moved up to 40th. In fairness, some argue that it should be higher when adjusting for the cost of living in the state.

[2]Roughly 92 percent of school funding comes from state and local sources, almost perfectly split between them (∼46 percent each) However, Oklahoma is a bit of an outlier with schools receiving nearly 70 percent of their budget from the state.

[3]See Barber (2022) for the history and implementation of four-day school weeks.

changes in salaries than changes in pensions and that shifting retirement benefits forward in time as salary is greatly preferred by teachers, but Oklahoma has employed a defined-benefit compensation structure that is much more weighted towards retirement than other similarly-salaried states.[4] Although, like most of the defined-benefit retirement plans that are still in existence, the state has practiced fiscal restraint in recent years and has made multiple changes to the rules for retirement that encourage longer tenure, require longer vesting periods, and actually incentivize coming out of retirement and returning to the classroom.[5] [6]

The state's creativity has not been limited to removing work days and tempting retired educators to come back to teach. To combat the difficulty many districts have had finding qualified teachers to fill classrooms, the state has tried to reduce as many barriers to entry as possible. Like numerous other states, Oklahoma allows for the use of an *emergency certificate* to be issued in instances when districts demonstrate an inability to hire qualified teachers.[7] Although these were legally available to all districts, not a single one was used before 2011, but the 2021-2022 school year will see nearly 4,000 emergency credentialed teachers in the classroom (Barber, 2022).[8] In addition, the state has recently eliminated the Oklahoma General Education Test (OGET), a required test to teach in the state, likely in response to declining pass rates in the years leading up

[4]While Oklahoma is ranked in the bottom quintile in salary, it falls into the third quintile in both average and median annual pension benefits.

[5]2022 Oklahoma Teacher Retirement System Handbook

[6]oksenate.gov/press-releases/bill-attract-retired-teachers-back-classroom-passes-committee

[7]"Qualified teacher" in this setting means a teacher holding the appropriate subject-area teaching certificate for their teaching assignment, e.g., early childhood education.

[8]This is out of approximately 41,000 teachers.

to the test's elimination.[9]

Looking further upstream, it is also a function of the supply of potential teachers graduating from college within the state, colloquially known as the "teacher pipeline", which has also been on the decline in recent years. According to the Oklahoma State Regents for Higher Education, the number of students graduating with a bachelor's degree in education has decreased 33 percent between 2011 and 2021 with the majority of that occurring in the last five years.[10]

In sum, there are a number of competing forces affecting the potential supply of teachers. This paper will investigate the trends in teacher retention, the variables affecting teacher tenure, and the response to the adoption of four-day weeks.

3.2 Data

This paper uses the same employee-panel data from the Oklahoma Department of Education (OSDE) that was used in prior research (Barber, 2022). This panel spans the school-years 2006-2007 to 2021-2022. However, we cannot use all of the data because of the nature of survival analysis. More specifically, our data is both left- and right-censored. On the left side, we are unable to observe how long teachers had been at their schools/districts in the first year of the panel. While we do have *total experience* as a variable for all teachers and all years, this is only weakly correlated with tenure (length of spell) within a given school or district. The obvious exception to this would be the

[9] www.okcu.edu/artsci/departments/education/certification-examination-pass-rates
[10] www.okhighered.org/studies-reports/outcomes/Degrees%20Granted/2020-21-report.pdf

subset of teachers for whom 2006 is their first year teaching (zero prior experience). Therefore, we remove any teaching spells that start before 2007. On the other side of the panel, we have right-censored data because we don't see time to failure (read: moving or leaving) for all teachers. This is not a problem and is, in fact, the norm for for survival analysis, but because our outcome(s) of interest require looking into the future to see each teacher's employment status, we have to drop the final year, 2021, from our dataset, but this data is used to inform the prior year's attrition before it is dropped. In fact, having censored data is one of the primary arguments for using a hazard model because of the problems that this presents for use of a linear probability model (Jenkins, 2005).

The employment panel includes the name of the school, district, and county for each teacher-year, along with each teacher's first and last name, race, gender, total experience teaching, grade(s) taught, subject area, full-time equivalents worked (FTEs) (which range from 0 to 1, with 1 equaling "full-time"), position title, and salary. These salaries are all transformed to 2021 dollars. We also have an indicator for each teacher's retirement status since some teachers return to the classroom after retiring. These teachers had salary caps placed upon them for most of the years in the sample, so they are examined separately.[11] Looking one year ahead, I am able to generate a binary indicator variable for each teacher-year that describes whether or not the teacher leaves the school, leaves the district, or leaves the profession in the following year. These

[11] Oklahoma, like other states with defined-benefit pensions, restricted the amount retirees could earn (in addition to their pension) if they returned to the classroom post-retirement to a maximum of $15,000. This cap was temporarily lifted in July 2017 and has been recently renewed until 2024 (SB 267).

forward-looking indicators are coded for the prior year ; i.e., a teacher that appears in the panel in 2017 but not 2018 will be indicated as having left teaching after the 2017 school year.

Unfortunately, what this data does **not** have is a consistent identifier for teachers, as described in Barber (2022), and teachers were linked using a series of author-written algorithms.

In addition, the employment panel data is merged with the National Center for Education Statistics (NCES) annual Common Core of Data (CCD) for demographic information about each school. This data set includes the racial makeup of each school, enrollment counts, percentage of students receiving free lunch, and number of full-time equivalent employees (FTEs) in addition to information about school location (address, latitude/longitude, and locale type), and Title I status.[12] Additional controls from the American Community Survey are used to capture local economic conditions, which are oftentimes important to teacher labor supply decisions.

Finally, I also use the panel of four-day school week adoption assembled and used by Thompson (2021) and Barber (2022) which spans the employment panel.

Our unit of observation is a *teaching spell*, which is simply defined as an unbroken period of employment within a given district. As mentioned earlier, at the end of each spell, each can be categorized as having ended to censoring (we are unable to see the end of the teaching spell), transfer (the teacher continues teaching but at another

[12]Title I is a federal program that subsidizes schools that are lower SES. For more details on the history of the program and its estimated impacts, see Matsudaira, Hosek, and Walsh (2012) and Cascio and Reber (2013).

district), or quitting/leaving (they do not appear in the panel in the following year).

3.3 Empirical Strategy

3.3.1 Record Linking

While many states make their employment records publicly available, there is no requirement for the inclusion of a unique identifier that would allow the longitudinal linking of employees across time. Naturally, this creates a problem for those studying determinants of attrition and would typically require the use of a "fuzzy matching" algorithm (Enamorado, Fifield, and Imai, 2019). Fortunately, Oklahoma is one of the very few states that include identifiers in their publicly-available employment records. However, the system used to track teachers has two significant problems with it. First, it changes twice during the analysis period. While this is less than ideal, a shifting, inconsistent identifier is much more helpful than none at all. Second, teachers are sometimes given different identifiers over time if their tenure is not continuous.

3.3.2 Duration Model

This paper employs a duration (also known as a hazard or survival) model to examine determinants of teacher tenure. These models estimate the conditional probability that a teacher leaves his or her district, given that he or she has not left in the prior year. Duration models are frequently used to study teacher attrition (Adams and Dial, 1993; Imazeki, 2005; Vagi, Pivovarova, and Miedel Barnard, 2019). These models are often preferred over ordinary least squares (or a binary dependent variable regression model) for modeling survival because of their ability to better handle censored data (than OLS) and to better model time-to-event, i.e., length of teaching spell (than logit or probit) (Jenkins, 2005).

Wu and Wen (2022) even provide us a heuristic for determining which model is more appropriate:

> If the binary outcome is best viewed as something akin to a biased coin flip, then a linear probability [difference-in-differences] may well be appropriate. But if the binary outcome is best viewed as a single-decrement, continuous-time process involving the transition from one discrete state to another, then the linear probability [difference-in-differences] should be avoided and a hazard [model] used instead. For some, this conclusion and our formal results may be seen as the unsurprising consequence of model misspecification. Still, that a binary outcome generated by a hazard process differs fundamentally from a biased coin flip – something long understood in the field of demography – is perhaps less well recognized, at least by some in other disciplines.

Usage of a duration model is also often preferred because of its semiparametric, flexible specification and, in the case of the Cox proportional hazard specific model described below and used in this paper, doesn't require any assumptions about the functional form of the underlying baseline hazard function. This can be seen in Equation 3.1

$$\theta(t, X) = \theta_0(t) \exp\left(\beta' X\right) = \theta_0(t)\lambda \tag{3.1}$$

$\theta(t, X)$ is a separable function comprised of two parts:

First, $\theta_0(t)$ is the "baseline hazard function", which only depends on t and not on the covariates contained within X. It summarizes the pattern of "duration dependence", which is assumed to be common to all teachers. This is a strong assumption that can be relaxed by separating individuals into different "strata" if they have sufficiently different attrition patterns, like male and female teachers, for example, which allows these different groups to have different baseline hazards while maintaining one set of coefficients (Kleinbaum, 1996).[13]

Second, $\lambda = \exp(\beta' X)$ is a teacher-specific function of time-invariant covariates, X, which scales the baseline hazard function common to all teachers. This is known as the "proportional hazards assumption" because it implies that the impact of a change in any variable contained in X affects the baseline hazard function of each stratum proportionally, independent of time. This observation (Equation 3.2) was first made by David Cox, so this model is commonly referred to as the *Cox model*.

$$\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)} = \exp[\beta_k (X_{ik} - X_{jk})] \tag{3.2}$$

Furthermore, if we assume only a one-unit change in X_k, ceteris paribus, then

$$\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)} = \exp(\beta_k) \tag{3.3}$$

The right side of this expression is known as the *hazard ratio*. The interpretations of these coefficients is relatively straightforward: values less than 1 imply that this variable is associated with longer tenure, and values greater than 1 imply reduced

[13]The use of strata can be seen as functionally equivalent to using fixed effects in OLS.

tenure, all else equal. The interpretation of these hazard ratios is also easy to interpret by converting it to a "percentage reduction in risk"

$$(1 - \text{Hazard Ratio}) \times 100\% \tag{3.4}$$

The Cox model has evolved to now permit time-varying covariates, so I am able to include time-varying teacher, school, district, and county-level controls. The inclusion of any covariate requires the testing to determine whether or not the proportional hazards assumption has been violated. For any models that had included covariates that violated the proportional hazards assumption, these models were stratified by those variables and it is indicated below the model table.

3.4 Results

Given the previous discussion of the regionally-idiosyncratic nature of the labor market, we begin with a descriptive analysis of the Oklahoma's labor market for teachers. Our dataset allows us to observe both the distribution of experience levels in addition to duration of employment within a given district (length of teaching spell). Figure 3.1 displays the distribution of experience levels of teachers at both the very beginning (2006) and near the end of our sample.[14] The differences are not substantial, but still noticeable. To begin with, 2019 actually had the highest percentage of teachers

[14] The pandemic likely distorted this distribution in a variety of ways: reducing attrition for the reduced costs of remote learning (far fewer new hires in 2021), accelerating some retirements, or in extreme cases, causing premature deaths. For this reason, 2019 is chosen as the last year.

with zero prior experience in the entire sample.[15] This, of course, squares with the documented increase in usage of emergency certificates for teachers; schools are hiring lots of new teachers, more than they've had to in the recent past, but the market is not functioning insofar as schools cannot find teachers with the needed credentials to teach.

Next, there are differences in the distribution of teachers in each experience bin in the early years of tenure, but this is likely a kind of ripple effect from the hiring trends of previous economic cycles, combined with changes in attitudes about teaching during the intervening years. For example, the unexpected probability mass for 5-9 years of experience would imply that these teachers starting teaching between 1997-2001, a time period marked by the lowest rate of teacher turnover in the last 25 years (Carver-Thomas and Darling-Hammond, 2019). This low turnover would further imply that these teachers had fewer outside options, and therefore got through the initial years in which teacher turnover is typically quite high. But, to a large degree, barring major changes to these variables, 2019's distribution should roughly be a horizontal shift of 2006's CDF.

Lastly, we can see some signs of the effect that recent changes in retirement rules have had on both retaining older, more experienced teachers and bringing formerly retired teachers back into the classroom. Prior to November 1, 2011, teachers hired before July 1, 1992 became eligible for full retirement benefits when either of two conditions were met: (a) the individual had reached age 62 or (b) the sum of age and years of service was greater than or equal to 80. For someone beginning teaching at 22

[15] 3,128 out of 40,584 full-time teachers.

and who teaches continuously, this would imply retirement eligibility after 29 years of service.[16] Indeed, it is at this age (level of experience) that we begin to see a drop-off in the number of teachers. After November 1, 2011, the "rule of 80" was changed to the "rule of 90", which meant that Condition (b) now required the sum of age and years of service to equal 90 for those hired after July 1, 1992. Using the example of the hypothetical teacher just mentioned, this would imply an additional five years of service to reach full benefits, which contributes to the growth in the tail between 2006 and 2019. In addition, Oklahoma has recently lifted salary caps for returning teachers who were already receiving retirement benefits, and this has also greatly impacted the experience distribution. During a period where the total number of teachers was relatively unchanged, the number of teachers with more than 35 years of experience has increased 250 percent between 2006 and 2021 and the number of retired teachers that have returned to the classroom has increased nearly 290 percent.[17]

Shifting to early-career teacher behavior, the literature has historically been focused on the first five years, and for obvious reasons – over half of teachers leave the profession during this window. Figure 3.2 describes the distribution of spell lengths observed between 2006 and 2021 where this early-career attrition can be seen. Because the employment panel spans 15 years, we can observe how five-year retention has evolved over time, and this is shown for four different cohorts in Figure 3.3. While there are year-to-year fluctuations, one thing that is readily apparent is the increased rate of

[16] 80-22=58, 58/2=29 (22+29)+29=80.

[17] >35 years of experience: 416 (2006) to 1,037 (2021), retired teachers: 253 (2006) to 731 (2021)

attrition in early career teachers over time. Figure 3.4 provides even more granularity by looking at three-year retention for twelve different cohorts of teachers that begin teaching during the years spanned by the panel.

Regarding the impact of the four-day school week on teacher labor supply, Barber (2022) provides some evidence that adoption of the policy was associated with a reduction in attrition in the years following the schedule change. More specifically, the policy appears to have reduced the rate of attrition for teachers with no prior experience while also allowing schools to recruit more experienced teachers, who are less likely to attrit by virtue of their tenure, at the margin. We can also estimate these effects with using an alternate methodology – the Cox model.

To begin with, we can examine the effect that this schedule change had on teachers who were already employed (but started after 2006) at the school prior to adoption and, therefore, could not have selected into the policy. The results of this model are found in Table 3.1. Despite starting with a large number of teachers, the sample is diminished greatly when we restrict it to teachers who (a) start after 2006, (b) teach at eventual four-day schools, and (c) receive tenure prior to the schedule change. After conditioning on all of these requirements, we are quite underpowered in our ability to make conclusive claims about the effects of the schedule change on previously tenured teachers. Our fully specified Cox model (Model 3) is suggestive of a 13.3 percent reduction in attrition for these teachers, but it is not statistically significant, so we cannot rule out the possibility of no effect since the 95 percent confidence interval of our treatment effect spans a 34 percent reduction to a 15 percent increase.

Next, we can observe the effect this policy has on teachers who were also hired before the policy but who have not yet received tenure. This restricts our sample even further, but this is an especially important group of teachers to track because of how much attrition is concentrated in the first few (pre-tenure) years of teaching. Table 3.2 contains the results of this model. Once again, we are constrained by our sample size and restrictions and cannot make any claims of an effect of the policy on this group of teachers.

Lastly, we can revisit the evidence of Barber (2022) that suggested that a primary benefit of the four-day school week was a reduction in attrition of teachers with no previous teaching experience. Using our Cox model, we can look at the attrition behavior of these zero-experience teachers, but also at those teachers who have prior experience but begin teaching at treated schools **after** the schedule change. Tables 3.3 and 3.4 display the results of our model for these two groups of teachers, respectively. In the former model, which looks at brand new teachers, we are able to confirm these previous results, finding a statistically significant 51.3 percent reduction in attrition for this group. In addition, we also find a statistically significant reduction in attrition of 39 percent for new teachers who arrive at treated schools with prior experience. A visual of the survival curves in which these effects manifest for cohorts starting at treated schools before, during, and after treatment can be found in Appendix Figure A6.

3.5 Conclusion

This paper uses Oklahoma public schoolteacher employment data to make two contributions to the literature. First, it links teachers across a 15-year panel in order to describe the evolution in the labor market for teachers through different economic climates, changes in pay, and even the elimination of a school day. We observe that early-career teachers are attriting at higher rates than ever before, but older teachers are retiring at much older ages than in years past, and this is likely due to changes in the rules regarding pension eligibility and the elimination in salary caps for retired teachers that continue to teach. Second, it is the first paper to quantify the impact of the four-day school week on teacher labor supply. While there has been anecdotal evidence that teachers have a revealed preference for a shorter work week, it had not yet been quantified, nor had the heterogeneous effects on existing versus new or experienced versus inexperienced. I find evidence suggestive of minimal effects to existing teachers already teaching at adopting schools before adoption. In addition, I find strong evidence of a large impact on teachers that select into the policy, both those with and without prior experience, but the effect is likely larger in teachers brand new to teaching. These retention effects, along with the recruitment effects observed in Barber (2022), are becoming the dominant reasons for districts to adopt the schedule. Once defended on fiscal responsibility grounds, the four-day school week is now being used as a tool used to attract teachers to schools that struggle to receive more than one application for posted vacancies.

Like many crises, the impending "teacher shortage" has been foretold for decades – it was first written about 100 years ago, experienced a resurgence in the 1950s and 1960s, and is once again gaining attention today.[18] And just like other crises, it is not clear when the prediction has come, or will come, to fruition. Also, how does one claim a "shortage" when one is talking about the occupation that employs more women than any other in the United States?[19] A logical place to begin would be to look upstream and examine the state of the teacher pipeline. As previously mentioned, the number of education degrees has been declining since the 1970s, but just like the teacher labor market, this is also subject to regional variation. In Oklahoma this year, multiple colleges have decided to cancel their education certification programs because of declining enrollments.[20] This is consistent with Sutcher, Darling-Hammond, and Carver-Thomas (2019) who modeled the supply and demand of teachers across time and found that demand for teachers began outpacing supply in 2013, the year in which Oklahoma began ramping up its utilization of emergency credentials – one of the clearest signs that schools cannot find qualified teachers to put in classrooms.

Districts are, of course, not completely to blame. First, the salaries are largely set by the state, independent of what might be necessary to clear the market, and this is especially true for Oklahoma where the large majority of funding comes from the state. Second, the cost of college tuition is increasing as teaching salaries are declining, both relative to other professions requiring a college degree and in real terms. In addition,

[18] books.google.com/ngrams/graph?content=%22teacher+shortage%22&year_start=1900&year_end=2019
[19] www.dol.gov/agencies/wb/data/occupations/most-common-occupations-women-labor-force
[20] www.news9.com/story/61d73c94ce36750be469dcf9/dipping-enrollment-leads-to-suspensions-of-education-preparation-programs

teaching salary increases for years of service are quite modest, and much of the compensation is back-loaded in the form of a defined benefit pension. However, even these once-generous pensions have become much less so as states have become increasingly concerned about their sustainability and are now requiring delayed retirement (both age and years of service) in order to maximize one's pension. All of these factors have led to the current situation: a race-to-the-bottom across the country in which states are removing as many barriers to teaching as possible while showing little regard for the effect that these changes may have on student achievement.

These shortages, like many economic phenomena in recent years, have hit rural areas particularly hard. It is, therefore, no surprise that these areas continue to be the ones that are most willing to experiment with such a drastic schedule change, especially considering the results in this paper. With the suggestive evidence that adoption of the four-day school week has minimal effects on existing faculty and conclusive evidence that the schedule retains new hires at significantly higher rates, these findings imply large benefits to the district, and at an estimated cost of $20,000-$30,000 per teacher needing replacement, reducing teacher turnover offers additional monetary dividends to cash-strapped districts on top of the explicit, direct benefit of the compensating differential of a four-day work week.[21] At the very least, these benefits must be weighed against the student achievement costs found in Thompson (2021). Using the value-added framework from Chetty, Friedman, and Rockoff (2014), Thompson finds a present-value reduction

[21] See Watlington et al. (2010) and Carver-Thomas and Darling-Hammond (2019) for the methodology behind these estimates.

in lifetime earnings of $2,140 per student per year of exposure to the four-day school week or more than $420,000 for the median four-day district. Using SEDA district test score averages, Thompson estimates less severe losses (one-half to one-quarter) for Oklahoma. Thus, for a similarly-sized school, any benefits derived from the schedule change would need to be on the order of $100,000 to $200,000 per year to mitigate the costs absorbed by the students[22]. Since we have demonstrated that existing, tenured teachers don't seem to be adversely affected by the change, and the majority of teachers ($\sim$ 75 percent) are tenured, it seems unlikely that reduced turnover costs alone would be enough. A potential future research topic would be to estimate the compensating differentials experienced by new hires at four-day schools – it is quite possible that these amounts could go quite far in mitigating any achievement costs, and eliminating one school day each week might actually end up being one of the more reasonable approaches that states employ if both education funds and teacher pipelines runs dry.

[22]This is putting aside any discussion of whether or not this is a transfer that should even be considered, but schools are clearly making this tradeoff.

References

Adams, Gerald J and Micah Dial (1993). "Teacher survival: A Cox regression model". In: *Education and Urban Society* 26.1, pp. 90–99.

Barber, Andrew (2022). "Essays on Effects of Public Policies". PhD thesis. UC Santa Cruz.

Biasi, Barbara (2019). "Higher Salaries or Higher Pensions? Inferring Preferences from Teachers' Retirement Behavior". In: *Inferring Preferences from Teachers' Retirement Behavior (October 31, 2019)*.

Carver-Thomas, Desiree and Linda Darling-Hammond (2019). "The trouble with teacher turnover: How teacher attrition affects students and schools." In: *Education Policy Analysis Archives* 27.36.

Cascio, Elizabeth U and Sarah Reber (2013). "The poverty gap in school spending following the introduction of Title I". In: *American Economic Review* 103.3, pp. 423–27.

Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood". In: *American economic review* 104.9, pp. 2633–79.

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai (2019). "Using a probabilistic model to assist merging of large-scale administrative records". In: *American Political Science Review* 113.2, pp. 353–371.

Hanushek, Eric A, Marc Piopiunik, and Simon Wiederhold (2019). "The value of smarter teachers international evidence on teacher cognitive skills and student performance". In: *Journal of Human Resources* 54.4, pp. 857–899.

Imazeki, Jennifer (2005). "Teacher salaries and teacher attrition". In: *Economics of education Review* 24.4, pp. 431–449.

Jenkins, Stephen P (2005). "Survival analysis". In: *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK* 42, pp. 54–56.

Johnston, Andrew C (2021). "Preferences, Selection, and the Structure of Teacher Compensation". In.

Kleinbaum, David G (1996). "The stratified Cox procedure". In: *Survival Analysis.* Springer, pp. 171–210.

Matsudaira, Jordan D, Adrienne Hosek, and Elias Walsh (2012). "An integrated assessment of the effects of Title I on school behavior, resources, and student achievement". In: *Economics of Education Review* 31.3, pp. 1–14.

NEA Research (2022). *2022 NEA Rankings Estimates Report.* Tech. rep.

Oklahoma State Dept. of Education (2021). *Oklahoma Teacher Supply and Demand Report*. Tech. rep.

Sutcher, Leib, Linda Darling-Hammond, and Desiree Carver-Thomas (2019). “Understanding teacher shortages: An analysis of teacher supply and demand in the United States.” In: *Education Policy Analysis Archives* 27.35.

Thompson, Paul N (2021). “Is four less than five? Effects of four-day school weeks on student achievement in Oregon”. In: *Journal of public economics* 193, p. 104308.

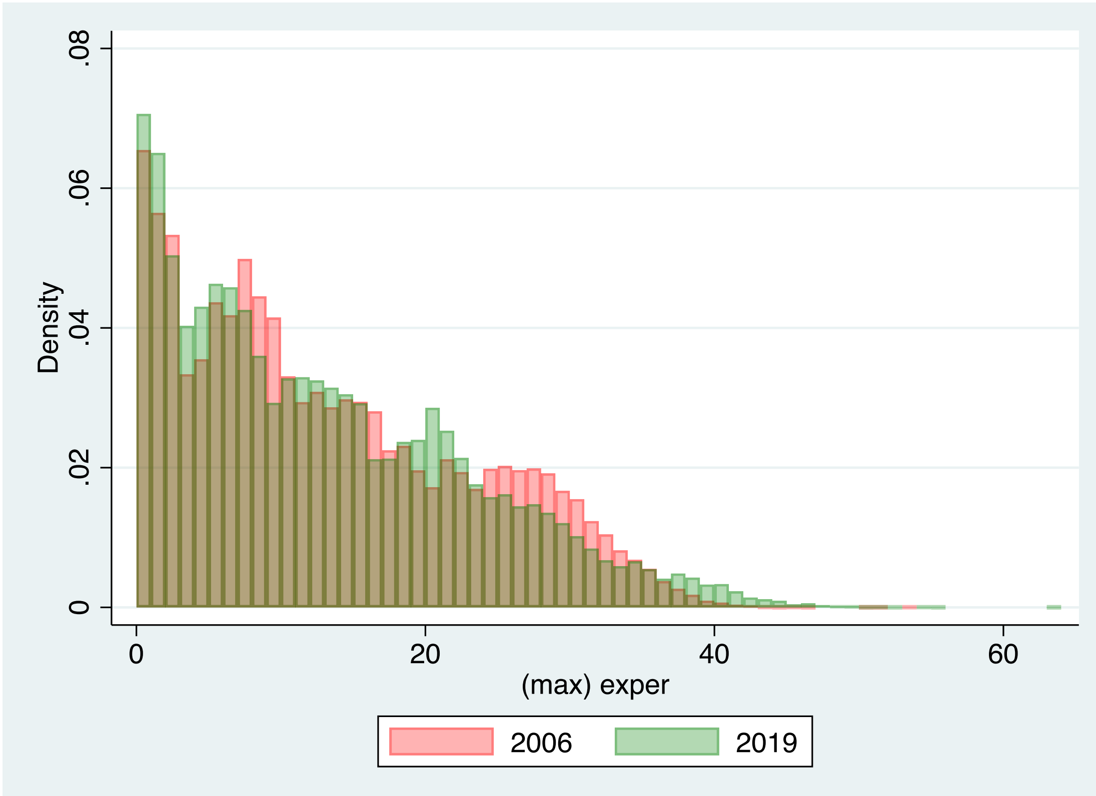
Vagi, Robert, Margarita Pivovarova, and Wendy Miedel Barnard (2019). “Keeping our best? A survival analysis examining a measure of preservice teacher quality and teacher attrition”. In: *Journal of Teacher Education* 70.2, pp. 115–127.

Watlington, Eliah et al. (2010). “The high cost of leaving: An analysis of the cost of teacher turnover”. In: *Journal of Education Finance*, pp. 22–37.

Wu, Lawrence L and Fangqi Wen (2022). “Hazard Versus Linear Probability Difference-in-Differences Estimators for Demographic Processes”. In: *Demography* 59.5, pp. 1911–1928.

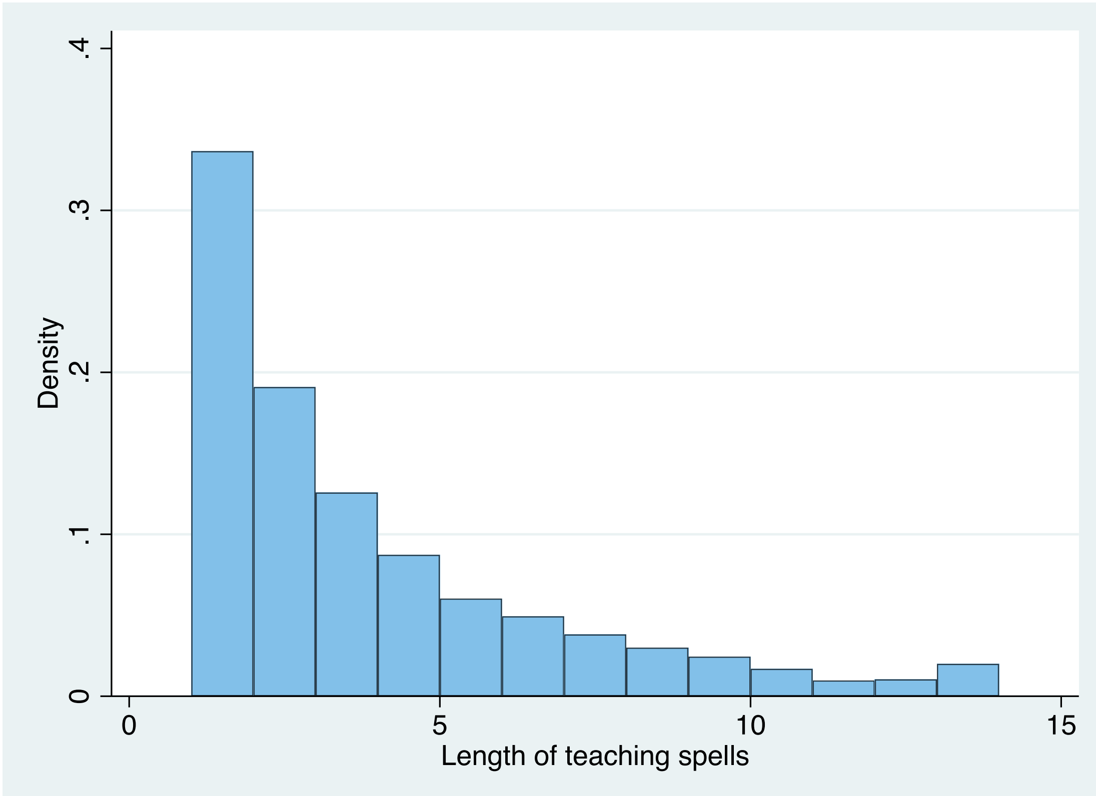
3.7 Figures and tables

Figure 3.1: Distribution of experience levels in 2006 and 2019



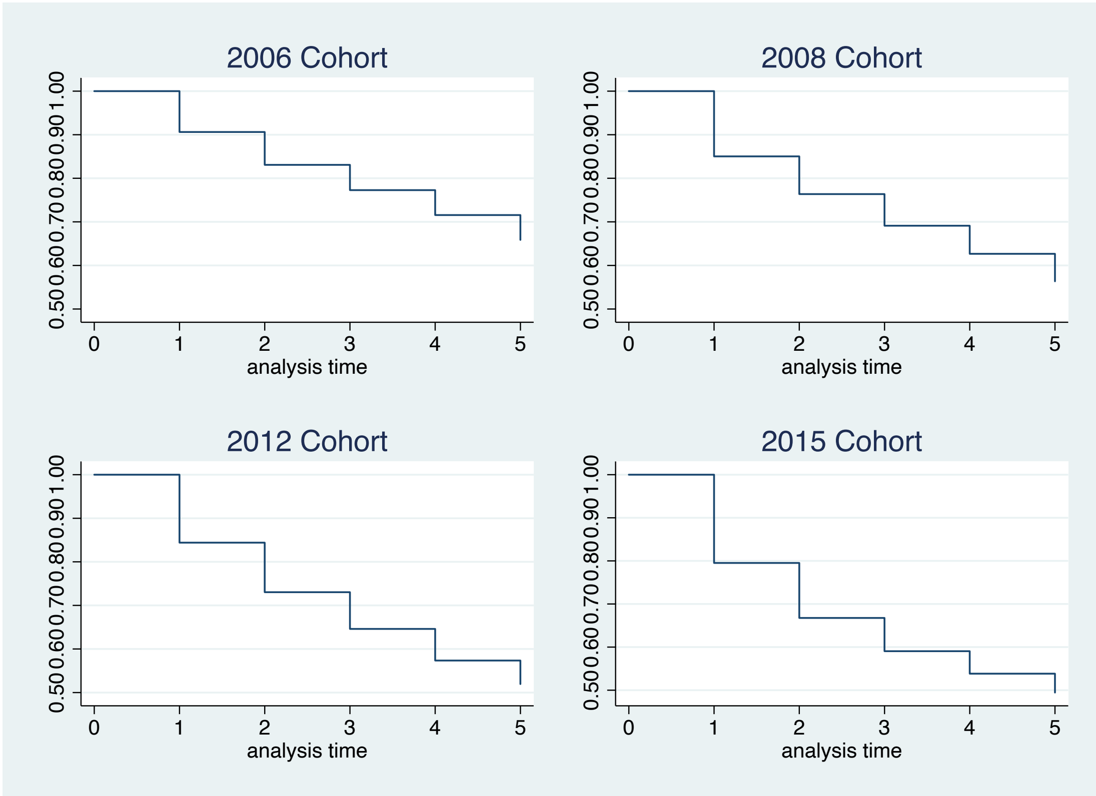
Notes: Distribution of experience levels across all full-time teachers in Oklahoma at the beginning of the 2006-2007 and 2019-2020 school years.

Figure 3.2: Distribution of experience levels upon leaving teaching



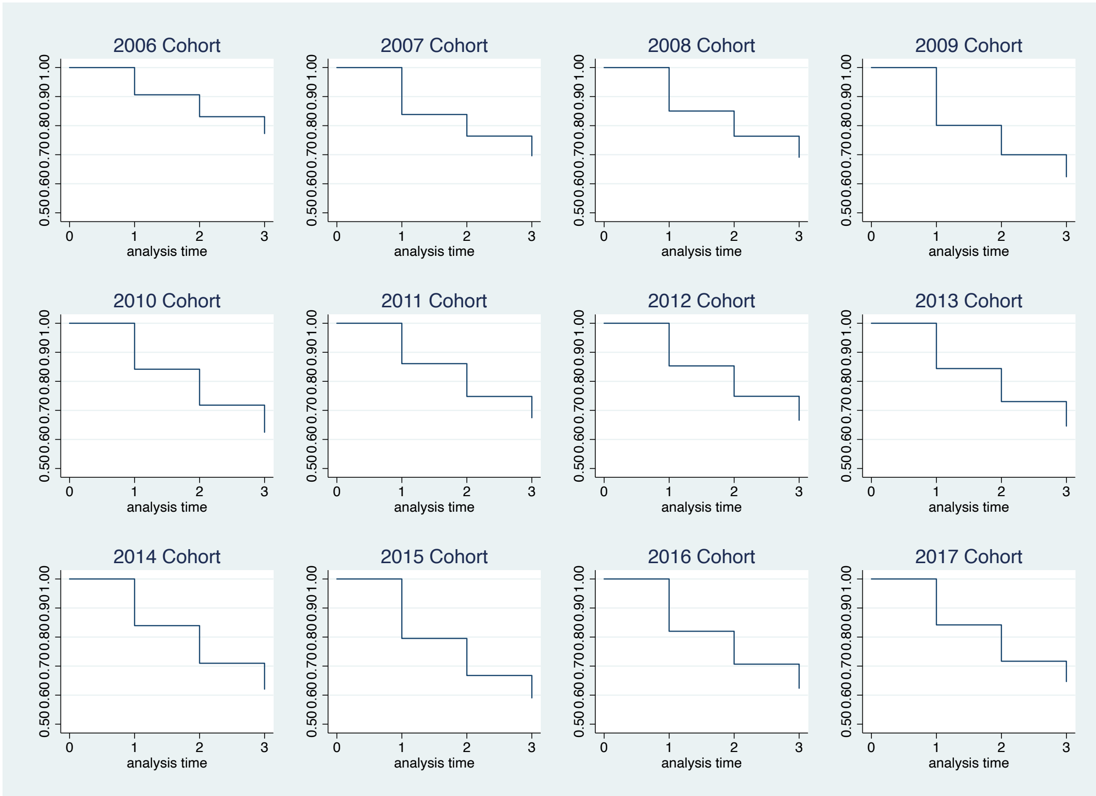
Notes: Distribution of spell lengths (unbroken periods of teaching in the same district) between the 2006-2007 and 2021-2022 school years.

Figure 3.3: Five-year retention rate of different cohorts



Notes: Five-year retention of teacher cohorts that began teaching in 2006, 2008, 2012, and 2015.

Figure 3.4: Three-year retention rate of different cohorts



Notes: Three-year retention of teacher cohorts who started teaching between 2006 and 2017.

Table 3.1: Cox Model: Effect of FDSWs on attrition of tenured teachers hired pre-policy

	(1)	(2)	(3)
Four-day week	0.878	0.887	0.867
	[0.669,1.154]	[0.678,1.161]	[0.655,1.147]
Female		0.901	0.902
		[0.763,1.064]	[0.756,1.075]
Log(salary)		0.165^{***}	0.171^{***}
		[0.0830,0.330]	[0.0890,0.330]
Graduate degree			1.767^{***}
			[1.495,2.088]
Math			1.040
			[0.819,1.321]
Science			0.902
			[0.673,1.210]
Frac. free lunch			0.620^{*}
			[0.353,1.088]
Observations	9212	9212	9186

Exponentiated coefficients; 95% confidence intervals in brackets

Stratification by: starting exper., locale, and spec. ed

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The teachers used in this model were

(1) teachers who started teaching after 2006

(2) teachers who taught at eventually treated schools

(3) teachers who had received tenure prior to treatment (i.e., they have taught for 3 years or more)

Table 3.2: Cox Model: Effect of FDSWs on attrition of untenured teachers hired pre-policy

	(1)	(2)	(3)
Four-day week	1.115	1.124	1.112
	[0.749,1.660]	[0.756,1.671]	[0.746,1.657]
Female		0.892	0.865
		[0.711,1.118]	[0.670,1.116]
Log(salary)		0.112**	0.0559***
		[0.0179,0.701]	[0.00688,0.454]
Graduate degree			1.690***
			[1.238,2.307]
Math			0.997
			[0.679,1.464]
Science			1.042
			[0.727,1.493]
Rural			1.337*
			[0.991,1.803]
Town			0.967
			[0.597,1.564]
Frac. free lunch			1.633
			[0.853,3.128]
Observations	2753	2753	2751

Exponentiated coefficients; 95% confidence intervals in brackets
Stratification by: starting exper. and spec. ed
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The teachers used in this model were

(1) teachers who started teaching after 2006

(2) teachers who taught at eventually treated schools

(3) teachers who hadn't yet received tenure prior to treatment (i.e., taught for ¡ 3 years)

Table 3.3: Cox Model: Effect of FDSWs on attrition of new teachers hired post-policy

	(1)	(2)	(3)
Four-day week	0.481***	0.491***	0.487***
	[0.302,0.767]	[0.312,0.773]	[0.310,0.765]
Female		1.130	1.144
		[0.797,1.602]	[0.796,1.643]
Log(salary)		0.0885**	0.0554***
		[0.0134,0.586]	[0.00902,0.341]
Graduate degree			1.272
			[0.846,1.912]
Math			1.113
			[0.695,1.783]
Science			0.784
			[0.471,1.304]
Special education			1.513
			[0.898,2.548]
Frac. free lunch			0.568**
			[0.323,1.000]
Observations	1488	1488	1488

Exponentiated coefficients; 95% confidence intervals in brackets

Stratification by: starting exper. and locale

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The teachers used in this model were

(1) teachers who started teaching after 2006

(2) teachers who started at already treated schools

(3) teachers with zero prior experience

Table 3.4: Cox Model: Effect of FDSWs on attrition of new teachers hired post-policy

	(1)	(2)	(3)
Four-day week	0.569**	0.580**	0.612**
	[0.368,0.880]	[0.379,0.890]	[0.401,0.934]
Female		1.043	1.004
		[0.802,1.357]	[0.771,1.308]
Log(salary)		0.285**	0.168***
		[0.103,0.785]	[0.0628,0.451]
Graduate degree			1.589***
			[1.249,2.020]
Math			0.903
			[0.657,1.243]
Science			0.868
			[0.588,1.279]
Special education			1.231
			[0.887,1.709]
Frac. free lunch			0.889
			[0.470,1.680]
Observations	3605	3605	3604

Exponentiated coefficients; 95% confidence intervals in brackets

Stratification by: starting exper. and locale

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The teachers used in this model were

(1) teachers who started teaching after 2006

(2) teachers who started at already treated schools

(3) teachers with prior experience

Appendix A

Appendix tables and figures

Table A1: State COVID-19 vaccination initiatives using conditional cash lottery incentives

Announcement date (in 2021)	State	Program name	Eligible vaccinations	Registration process	Largest prize ($)	Total prize value (approximate $)	Exclusive prizes	Drawing frequency
May 12	Ohio	Vax-A-Million	All	Opt-in	1,000,000	5,600,000	Yes	Weekly
May 20	Maryland	VaxCash	All	Auto	400,000	2,000,000	Yes	Daily
May 20	New York	Vax And Scratch	New	Opt-in	5,000,000	Unknown	No	Instant
May 21	Oregon	Take Your Shot	All	Auto	1,000,000	1,500,000	Yes	Once
May 24	Delaware	DEWins	New/All	Auto	302,000	5,000,000	Yes	Semi-weekly
May 25	Arkansas	Not named	New	Opt-in	1,000,000	2,000,000	No	Instant
May 25	Colorado	Comeback Cash	All	Auto	1,000,000	6,250,000	Yes	Weekly
May 27	California	Vax For The Win	New/All	Auto/Opt-in	1,500,000	116,500,000	Yes	Weekly
May 27	West Virginia	Do It For Babydog	All	Opt-in	1,588,000	10,000,000	Yes	Weekly
June 01	New Mexico	Vax 2 The Max	All	Opt-in	5,000,000	10,000,000	Yes	Weekly
June 03	Washington	Shot Of A Lifetime	All	Auto	1,000,000	2,400,000	Yes	Weekly
June 04	Kentucky	Shot At A Million	All	Opt-in	1,000,000	4,200,000	Yes	Monthly
June 10	North Carolina	Summer Cash	New/All	Auto	1,000,000	4,500,000	Yes	Bi-weekly
June 15	Massachusetts	VaxMillions	All	Opt-in	1,000,000	5,500,000	Yes	Weekly
June 17	Illinois	All In For The Win	All	Auto	1,000,000	10,000,000	Yes	Weekly
June 17	Louisiana	Shot At A Million	All	Opt-in	1,000,000	2,300,000	Yes	Weekly
June 17	Maine	Don't Miss Your Shot	All	Opt-in	896,809	896,809	Yes	Once
June 18	Nevada	Vax Nevada Days	All	Auto	1,000,000	5,000,000	Yes	Weekly
July 01	Michigan	Shot To Win	All	Opt-in	2,000,000	5,495,000	Yes	Bi-weekly
July 21	Missouri	MO VIP	All	Opt-in	10,000	9,000,000	Yes	Bi-weekly

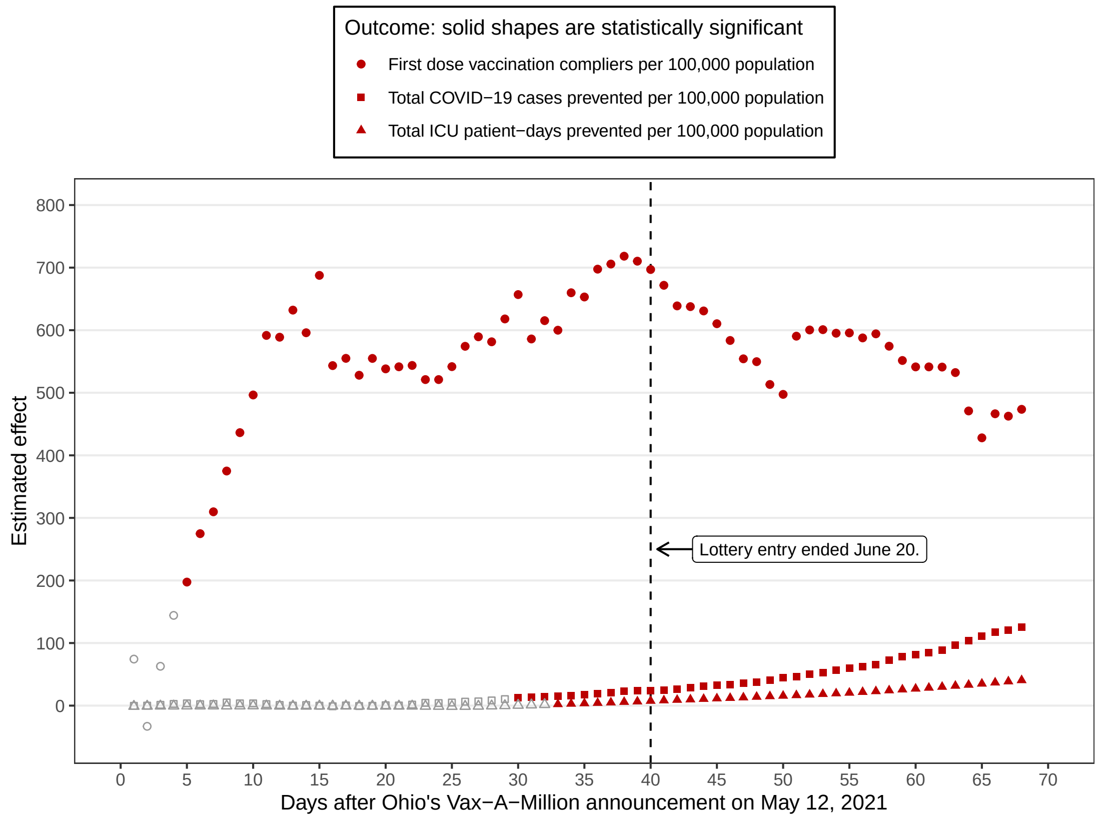
Notes: Table A1 lists all state-run conditional cash lottery incentive schemes for COVID-19 vaccinations in the United States. Ohio's Vax-A-Million incentive program was the first and was announced on May 12, 2021 and lottery entry ended on June 20, 2021. The eligible vaccinations column indicates whether people who were vaccinated prior to the program's announcement could win prizes in the CCL.

Table A2: State unit weights for the ridge augmented synthetic control models

State	Unit weights for model		
	Vaccinations	Cases	ICU days
Alabama	-0.065	-0.000	-0.000
Alaska	-0.021	-0.000	-0.000
Arizona	0.008	-0.000	-0.000
Connecticut	-0.058	-0.000	0.000
District of Columbia	0.009	0.000	-0.000
Florida	-0.067	-0.000	-0.000
Georgia	0.107	-0.000	0.000
Hawaii	-0.045	-0.000	-0.001
Idaho	0.181	0.000	0.000
Indiana	0.001	0.136	0.419
Iowa	-0.066	0.000	0.000
Kansas	0.281	0.304	0.082
Michigan	0.191	0.033	0.078
Minnesota	0.015	0.000	0.000
Mississippi	0.040	-0.000	-0.000
Missouri	-0.038	0.000	0.197
Montana	-0.119	-0.000	-0.000
Nebraska	0.115	0.000	0.000
New Hampshire	-0.044	0.000	-0.000
New Jersey	0.079	-0.000	0.060
North Dakota	0.126	0.000	0.000
Oklahoma	-0.007	0.000	0.000
Pennsylvania	-0.057	0.000	0.000
Rhode Island	0.024	0.109	0.165
South Carolina	0.043	-0.000	-0.000
South Dakota	-0.024	-0.000	0.000
Tennessee	0.049	0.000	0.000
Texas	-0.056	-0.000	0.000
Utah	0.091	0.141	0.000
Vermont	0.026	-0.000	-0.000
Virginia	0.031	0.000	0.000
Wisconsin	0.321	0.277	0.000
Wyoming	-0.067	-0.000	-0.000

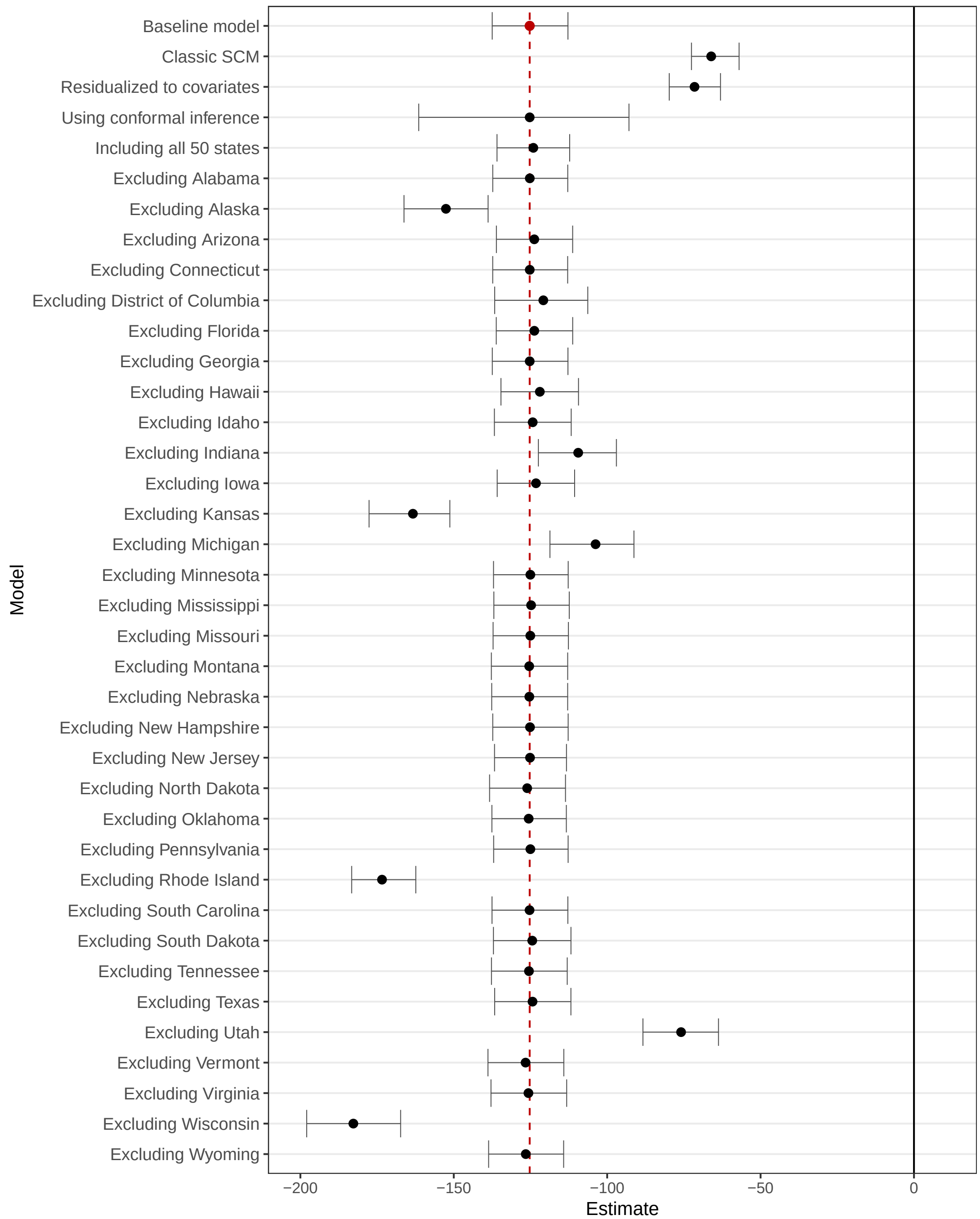
Notes: States not listed are not in the donor pool.

Figure A1: Estimated effects for transformations of each outcome into a common scale



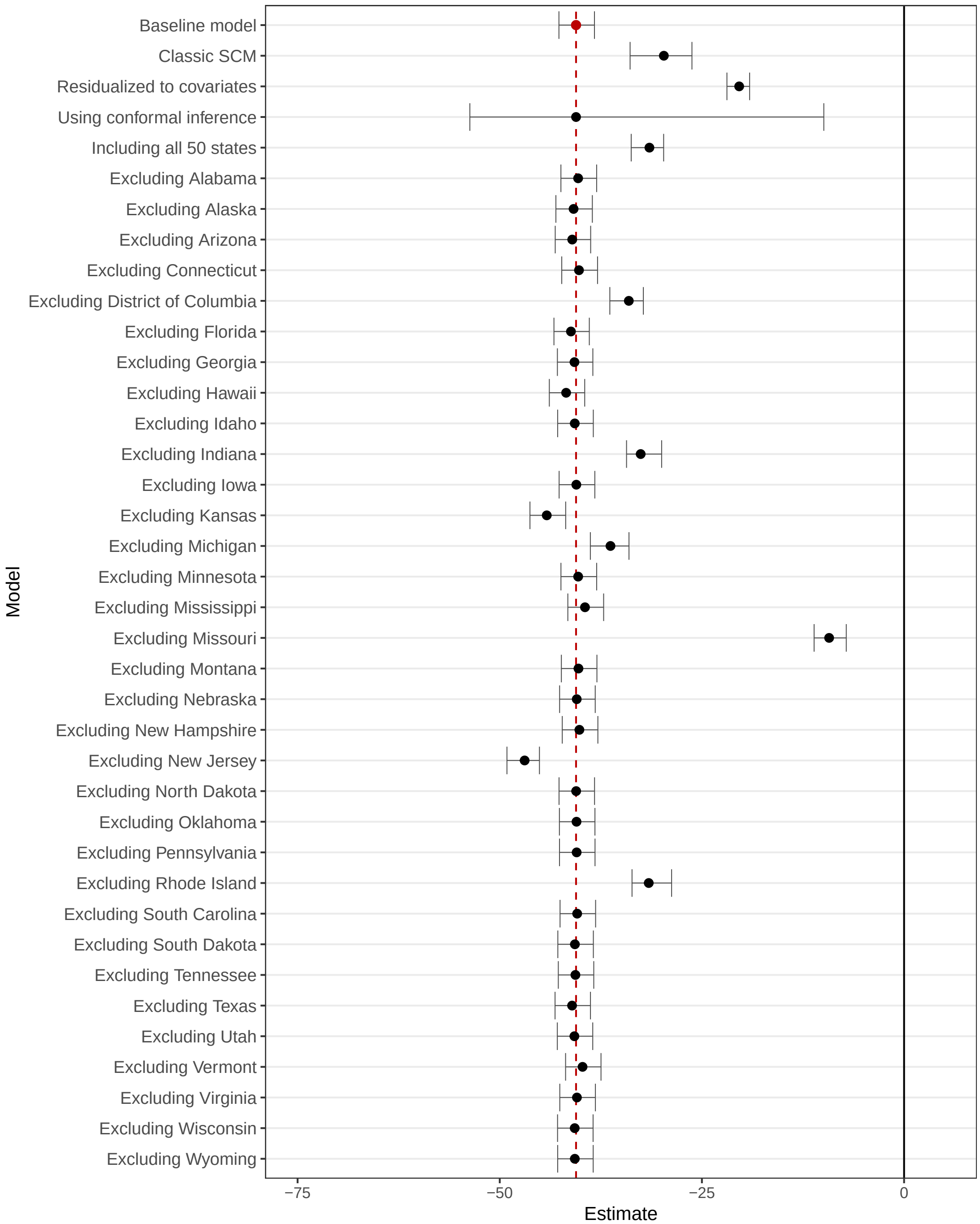
Notes: Figure A1 shows estimated differences between Ohio and the synthetic control for the three outcomes examined in this study, transformed to use a common scale. These transformations are: (1) The fraction of population with any COVID-19 vaccination – multiplied by 100,000. (2) The cumulative total COVID-19 cases recorded per 100,000 population – multiplied by negative one. (3) The cumulative total COVID-19 ICU patient-days per 100,000 population – multiplied by negative one. These effects are plotted by day following Ohio's Vax-A-Million lottery announcement. The shapes are solid if the 95 percent confidence interval does not overlap with zero, as calculated using jackknife+ inference.

Figure A2: Robustness checks of the synthetic control estimates for the cumulative total COVID-19 cases recorded per 100,000 population, using different samples and specifications



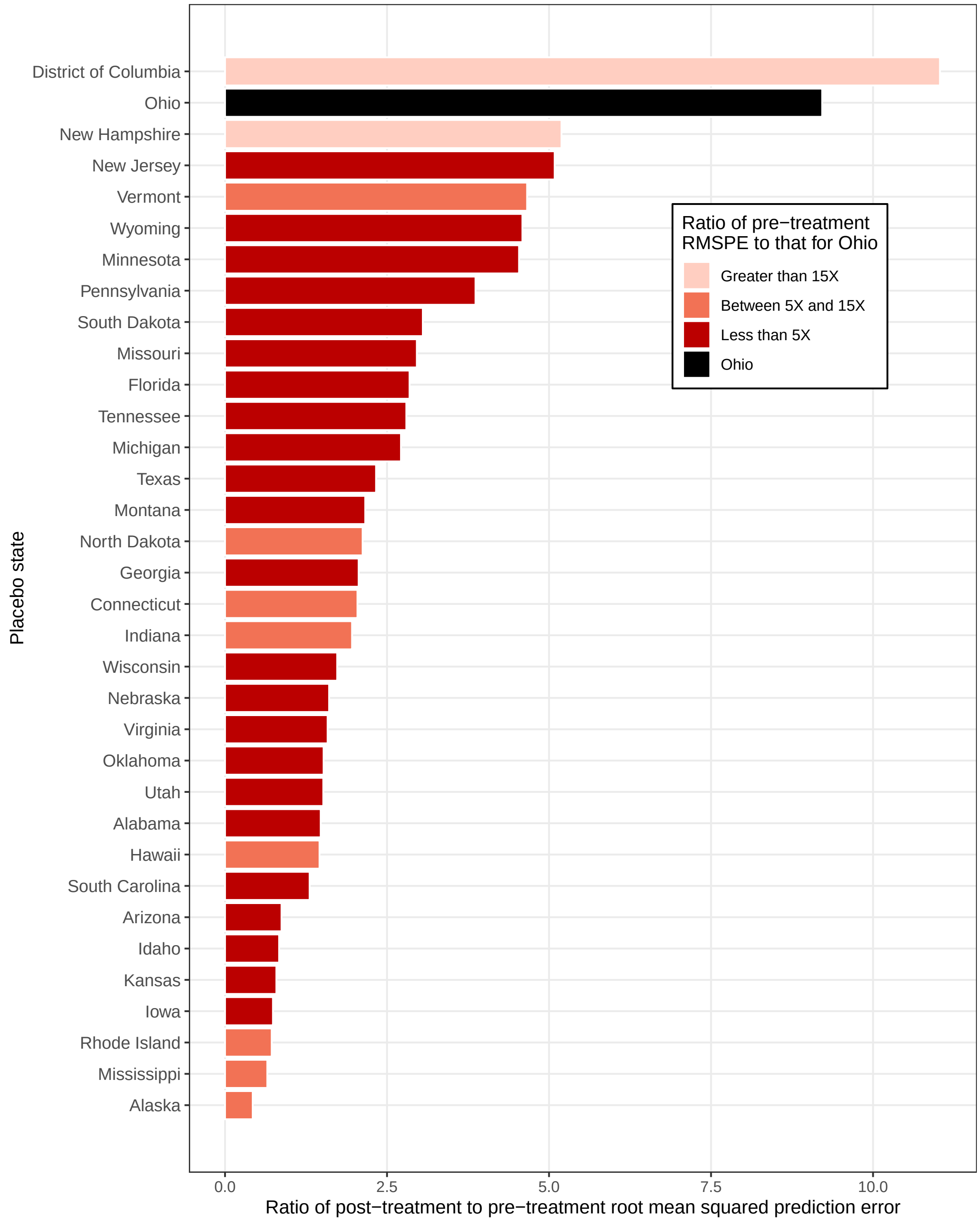
Notes: Figure A2 shows estimated differences between Ohio and the synthetic control for the cumulative total COVID-19 cases recorded per 100,000 population by July 18, 2021. Each row depicts results from a separate model using the data sample and/or specification denoted. The grey error bars indicate the respective 95 percent confidence intervals, which are calculated using jackknife+ inference except where indicated, i.e., the model that uses conformal inference.

Figure A3: Robustness checks of the synthetic control estimates for the total COVID-19 ICU patient-days per 100,000 population, using different samples and specifications



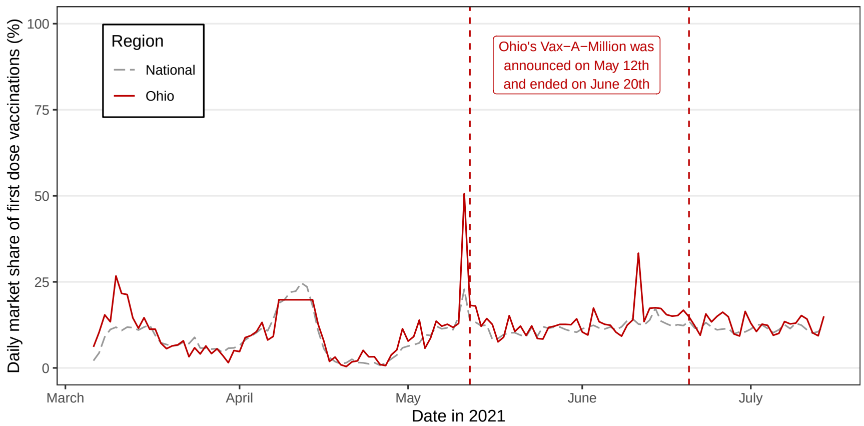
Notes: Figure A3 shows estimated differences between Ohio and the synthetic control for the cumulative total COVID-19 ICU patient-days per 100,000 population by July 18, 2021. Each row depicts results from a separate model using the data sample and/or specification denoted. The grey error bars indicate the respective 95 percent confidence intervals, which are calculated using jackknife+ inference except where indicated, i.e., the model that uses conformal inference.

Figure A4: Synthetic control placebo effects and rankings for other states

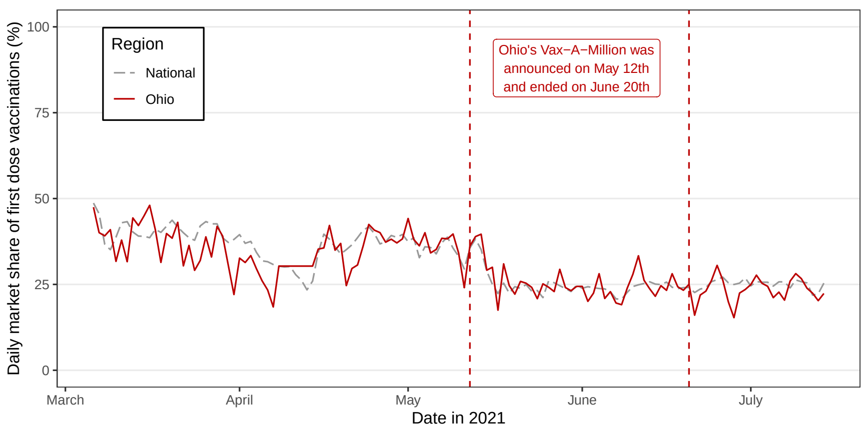


Notes: States not listed are not in the donor pool. The outcome is the share of population with any COVID-19 vaccination (at least a first dose). Post-treatment RMSPE are computed using the full treatment period, starting with Ohio's Vax-A-Million announcement on May 12, 2021 and ending with the lottery entry end-date on June 20, 2021. Pre-treatment RMSPE are computed using the full pre-treatment period in the data, starting on February 19, 2021 and ending on May 11, 2021.

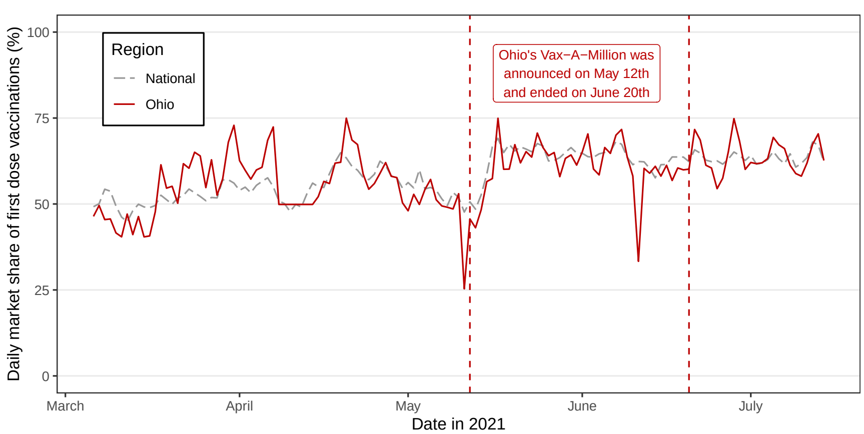
Figure A5: Manufacturers' daily market shares of first dose vaccinations over time



(a) Janssen



(b) Moderna



(c) Pfizer

Figure A6: Five-year retention of teacher cohorts at eventually treated schools