

# The Long Shadow of the Paddle? Evidence from State Corporal Punishment Bans

APEP Working Paper 34 nd @dakoyana

January 2026

## Abstract

We investigate whether state-level bans on corporal punishment in public schools affected long-term educational and economic outcomes. Using the staggered adoption of bans across 33 U.S. states between 1971 and 2023, we implement a difference-in-differences design with 3.2 million observations from the American Community Survey (2017–2022). Our event-study analysis reveals clear violations of the parallel trends assumption: pre-treatment coefficients show systematic differences between early-ban (predominantly Northeastern) and never-ban (predominantly Southern) states that predate any policy change. The counterintuitive finding that bans *increase* disability rates is a strong indicator of residual confounding from divergent regional trends rather than a causal effect. We conclude that the stark socioeconomic and cultural differences between states that adopted bans early versus never preclude credible causal identification with standard two-way fixed effects methods. Our results underscore the importance of pre-trends diagnostics (?) and careful counterfactual construction in staggered adoption designs, and motivate the use of synthetic control methods (??) when parallel trends cannot be maintained.

**Keywords:** corporal punishment, education policy, difference-in-differences, parallel trends, identification, school discipline

**JEL Codes:** I21, I28, C21, J24

# 1 Introduction

Corporal punishment—the use of physical force to discipline children—was nearly universal in American public schools until the 1970s. Teachers and administrators routinely employed paddling, spanking, and other forms of physical punishment to enforce classroom rules and maintain order. Despite growing psychological research documenting potential harms, the U.S. Supreme Court in *Ingraham v. Wright* (1977) held that the Eighth Amendment’s prohibition on cruel and unusual punishment did not apply to school discipline, leaving the legality of corporal punishment to individual states.

Beginning with Massachusetts in 1971, states began abolishing corporal punishment in public schools. By 2024, 33 states and the District of Columbia had enacted bans, though corporal punishment remains legal in 17 states—predominantly in the South—and is actively practiced in 12 (?). This staggered adoption across states over five decades provides potential quasi-experimental variation for estimating the causal effects of corporal punishment bans on long-term outcomes.

This paper asks: Do individuals whose K–12 schooling occurred entirely after their state banned corporal punishment have better educational attainment and economic outcomes as adults? The question has direct policy relevance: if bans improve long-term outcomes, they provide ammunition for abolition efforts in the remaining states. More broadly, understanding the effects of school discipline policy informs debates about classroom management, child development, and the appropriate boundaries of state authority over children.

We exploit the staggered timing of state bans to implement a difference-in-differences (DiD) design. Using data from the American Community Survey (ACS) on over 3.2 million adults born between 1955 and 1997, we compare outcomes for cohorts born in states that banned corporal punishment before their schooling began (“fully treated”) to cohorts born in states that never banned (“never treated”). Our primary outcomes include years of education, high school and college completion rates, employment, and disability status.

Our main finding is that **credible causal identification is not achievable** with this

research design. While our point estimates on education and employment are generally null or small, we also find counterintuitive results: a small negative effect on high school completion ( $-1.0$  percentage points,  $p = 0.047$ ) and a positive association with disability rates ( $+2.2$  percentage points,  $p < 0.001$ ). These anomalous findings are inconsistent with any plausible causal mechanism and instead signal residual confounding from divergent trends between the two groups of states. Our event-study analysis confirms the diagnosis: pre-treatment coefficients are not flat and zero as required by the parallel trends assumption, but instead show systematic pre-existing differences that reflect the fundamentally different educational trajectories of Northeastern early-ban states versus Southern never-ban states.

We interpret these results with caution. The states that adopted corporal punishment bans earliest—Massachusetts, Hawaii, Maine, and the northeastern states—differ systematically from the states that retained it—Texas, Mississippi, Alabama, and other Southern states. These differences include baseline educational quality, income levels, racial composition, and cultural attitudes toward discipline. Even with state and cohort fixed effects, our design cannot fully account for differential trends in outcomes across these groups of states over the 50-year period we study.

Our paper contributes to several literatures. First, we add to the economics literature on school discipline policy. ? demonstrate that exposure to disruptive peers—and by extension, the disciplinary responses to them—creates negative externalities for classmates. ? documents the persistent black-white gap in school discipline and its implications for educational inequality. While this literature has examined suspensions, expulsions, and restorative justice programs (?), corporal punishment has received less attention from economists despite its continued prevalence. Second, we contribute to the large psychology literature on corporal punishment by providing quasi-experimental evidence from a policy context, complementing the predominantly correlational evidence base (?). Third, and perhaps most importantly, we add to the methodological literature on staggered adoption designs by providing a cautionary case study of what happens when parallel trends fail. Our results illustrate the concerns

raised by ? and ? about the importance of rigorous pre-trends testing and the dangers of proceeding when identification assumptions are violated.

The remainder of the paper proceeds as follows. Section 2 provides background on corporal punishment policy and the psychological literature on its effects. Section 3 describes our data and empirical strategy. Section 4 presents results. Section 5 discusses limitations and threats to identification. Section 6 concludes.

## 2 Background

### 2.1 Corporal Punishment in American Schools

Corporal punishment has a long history in American education. Rooted in Puritan beliefs about the necessity of “breaking the will” of children, physical discipline was the norm in colonial schools and persisted throughout the 19th and 20th centuries. The practice typically involved paddling students on the buttocks with a wooden paddle, though other forms—slapping, spanking, and hitting with rulers—were also common.

By the mid-20th century, corporal punishment began facing opposition from educators and child psychologists who questioned its efficacy and morality. The 1960s and 1970s saw growing awareness of children’s rights and concerns about physical abuse. However, the Supreme Court’s 1977 decision in *Ingraham v. Wright* foreclosed federal constitutional challenges, holding that the Eighth Amendment applies only to convicted criminals, not schoolchildren.

In the absence of federal prohibition, reform occurred state by state. New Jersey was an early outlier, having banned school corporal punishment in 1867. The modern wave of abolition began with Massachusetts (1971), followed by Hawaii (1973), Maine (1975), Rhode Island (1977), and others throughout the 1980s and 1990s. The pace of reform has slowed dramatically: only Pennsylvania (2005), Ohio (2009), New Mexico (2011), and most recently Colorado and Idaho (2023) have enacted bans in the past two decades.

Today, 17 states still permit corporal punishment in public schools: Alabama, Arizona, Arkansas, Florida, Georgia, Indiana, Kansas, Kentucky, Louisiana, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Wyoming. The practice is concentrated in the South, where over three-quarters of all reported incidents occur in just four states: Alabama, Arkansas, Mississippi, and Texas (?).

## **2.2 Prevalence and Disparities**

When corporal punishment was at its peak in the 1970s, millions of students were physically disciplined each year. The practice declined substantially even in states that retained it, driven by changing norms, liability concerns, and local district-level bans. By 2014, approximately 109,000 students were subjected to corporal punishment nationwide—a fraction of historical levels but still substantial (?).

Research consistently documents racial disparities in corporal punishment administration. Black students receive corporal punishment at rates 1.4 to 2 times higher than white students, even after controlling for behavior (?). Male students are punished at much higher rates than female students. Students with disabilities also face elevated punishment rates. These disparities raise civil rights concerns that have motivated some district-level bans even in states where the practice remains legal.

## **2.3 Psychological Research on Effects**

A large body of psychological research examines the effects of corporal punishment on child development. The meta-analysis by ? synthesized 88 studies and found corporal punishment associated with 10 negative outcomes, including increased aggression, delinquency, mental health problems, and damaged parent-child relationships. Only one outcome showed a positive association: immediate compliance. However, even this benefit disappeared when examining long-term compliance.

Subsequent research has reinforced these findings. ? found that spanking was associated

with increased aggression even after controlling for initial levels of aggression, suggesting a causal pathway. Neuroimaging studies have documented structural brain changes associated with harsh physical punishment (?). The American Academy of Pediatrics, American Psychological Association, and virtually all major child welfare organizations now oppose corporal punishment.

However, the psychology literature faces methodological limitations. Most studies are correlational, comparing children who experience corporal punishment to those who do not. These comparisons may be confounded by parental characteristics, family environment, child behavior problems that precipitate punishment, and other factors. The few experimental studies involve parenting interventions rather than school policy changes.

Our study contributes quasi-experimental evidence by exploiting state-level policy variation. If the parallel trends assumption holds, comparing cohorts exposed to bans versus not exposed provides causal estimates free of individual-level confounding.

## **2.4 Theoretical Mechanisms**

Several theoretical pathways link corporal punishment bans to long-term educational and economic outcomes. Understanding these mechanisms is essential for interpreting any effects (or null findings) that emerge from our empirical analysis.

The most direct mechanism involves the reduction of trauma and psychological stress. Physical punishment causes immediate pain, fear, and humiliation for the students who receive it. For some students, particularly those subjected to repeated punishment, this trauma may interfere with learning by activating stress responses that impair memory formation and cognitive function (?). The fear of punishment may damage students' emotional attachment to school, reducing engagement and increasing the risk of truancy and dropout. By eliminating corporal punishment, bans could reduce school-related stress and create conditions more conducive to learning and development.

A second mechanism operates through school climate. Schools that rely heavily on phys-

ical punishment tend to have more authoritarian, adversarial disciplinary cultures that may damage student-teacher relationships and create environments of fear rather than trust (?). When corporal punishment is banned, schools must develop alternative disciplinary approaches such as counseling, restorative justice practices, and behavioral interventions. These alternatives may foster more supportive school environments that benefit all students, not just those who would have received physical punishment. The cultural shift away from violence as a disciplinary tool may have broader effects on how adults treat children in educational settings.

Third, corporal punishment bans may reduce absenteeism and school dropout. Students who have been subjected to physical punishment, or who fear future punishment, may avoid school entirely or disengage from education at the earliest opportunity. ? documented that students who experienced corporal punishment were more likely to report negative attitudes toward school and lower educational aspirations. By eliminating this source of fear and humiliation, bans could improve attendance rates and increase the likelihood that students complete high school.

Fourth, reduced exposure to violence in schools may improve mental health outcomes, which in turn could enhance educational attainment and labor market success. The psychological literature has documented strong associations between corporal punishment and depression, anxiety, and behavioral problems (?). If bans reduce these mental health burdens, affected cohorts may experience better outcomes across multiple life domains.

However, several countervailing mechanisms could attenuate or even reverse these positive effects. Most importantly, schools may substitute other disciplinary measures that are equally or more harmful than corporal punishment. The dramatic rise in school suspensions and expulsions since the 1970s—documented by ?—occurred precisely as corporal punishment declined. If suspended students lose instructional time, fall behind academically, and become disconnected from school, the net effect of banning corporal punishment while increasing suspensions could be negative. Similarly, increased referrals to school re-

source officers and the juvenile justice system could produce worse outcomes than corporal punishment would have.

A related concern is that eliminating corporal punishment could reduce deterrence of student misbehavior if alternative disciplinary measures are perceived as less threatening. Although the evidence for corporal punishment’s deterrent effect is weak (?), some educators and parents believe it provides immediate behavioral compliance. If bans lead to increased classroom disruption, all students’ learning could be affected, not just those who would have received punishment.

Finally, selection effects may confound any observed associations between bans and outcomes. States that chose to ban corporal punishment early may have done so because they already had stronger schools, more educated populations, and more resources to invest in alternative disciplinary approaches. These underlying characteristics, rather than the bans themselves, may drive any observed outcome differences. This selection concern motivates our empirical strategy, which attempts to difference out time-invariant state characteristics.

## **3 Data and Empirical Strategy**

### **3.1 Data Sources**

Our empirical analysis draws on two primary data sources: individual-level survey data from the American Community Survey and a comprehensive database of state corporal punishment ban dates that we compiled from legal and administrative records.

The American Community Survey (ACS) Public Use Microdata Samples (PUMS) provide the individual-level data for our analysis. The ACS is a large-scale, nationally representative survey conducted by the U.S. Census Bureau that collects detailed demographic, educational, and economic information from approximately 1% of the U.S. population annually. We use data from survey years 2017 through 2022, excluding 2020 when the Census Bureau did not release 1-year ACS estimates due to disruptions in data collection caused by the COVID-



19 pandemic. The resulting dataset contains rich information on educational attainment, employment status, disability status, and a range of demographic characteristics including age, sex, and race/ethnicity.

Crucially for our identification strategy, the ACS records each respondent’s state of birth. This allows us to link adults observed in the survey to the corporal punishment policy regime that prevailed in their birth state during their schooling years. We assume that individuals attended school in their state of birth—an assumption we discuss in our limitations section, as interstate migration introduces measurement error in treatment assignment.

We restrict our sample to U.S.-born adults between the ages of 25 and 64 at the time of the survey. The lower age bound ensures that respondents have completed their formal education, while the upper bound limits our sample to individuals young enough to have potentially been affected by corporal punishment bans enacted since 1971. Within this age range, we further restrict to individuals born in one of 16 selected states that provide substantial variation in the timing of corporal punishment bans.

Our state selection includes four early-ban states that abolished corporal punishment between 1971 and 1977: Massachusetts (1971), Hawaii (1973), Maine (1975), and Rhode Island (1977). These northeastern and Pacific states were pioneers in school discipline reform. We also include four mid-period ban states from 1985 to 1989: New York (1987), California (1986), Wisconsin (1988), and Minnesota (1989). Two late-ban states round out our treatment group: Washington (1993) and West Virginia (1994). Finally, we include six states that have never banned corporal punishment and continue to permit the practice: Texas, Florida, Georgia, Alabama, Mississippi, and North Carolina. These Southern states serve as our control group.

This selection of 16 states balances several considerations. First, it provides substantial variation in ban timing, spanning over two decades of policy adoption. Second, it includes both early adopters and never-adopters, enabling the comparison central to our difference-in-differences design. Third, it keeps the sample computationally manageable while retaining

sufficient statistical power. Our final analytic sample comprises 3,209,523 individual observations.

The second component of our data infrastructure is a comprehensive database of state-level corporal punishment ban dates. We compiled this database by reviewing state education codes, legislative histories, and administrative regulations for all 50 states and the District of Columbia. Primary sources included state statutory compilations, administrative code repositories, and state board of education records. We supplemented these primary sources with information from the academic literature, particularly the comprehensive policy reviews conducted by ? and ?. Where discrepancies arose between sources, we prioritized primary legal documents and verified dates against court decisions and contemporaneous news coverage when available. The resulting database records the year in which each state’s ban took effect, distinguishing between states that banned corporal punishment through legislation, administrative rule, or court order.

### 3.2 Variable Construction

The core of our empirical strategy is the assignment of individuals to treatment and control groups based on the timing of corporal punishment bans relative to their schooling years. We define treatment status using the relationship between an individual’s birth year, their state of birth, and the year in which that state banned corporal punishment (if ever).

For an individual born in state  $s$  in year  $y$ , we assume that they entered school at age 6 (in year  $y + 6$ ) and completed school at age 18 (in year  $y + 18$ ). If state  $s$  enacted a corporal punishment ban in year  $b$ , we classify the individual into one of three treatment categories. Individuals are considered “fully treated” if the ban occurred before they entered school, meaning  $b < y + 6$ . These individuals’ entire K–12 experience occurred under a policy regime that prohibited corporal punishment. Individuals are considered “never treated” if they were born in a state that has never banned corporal punishment, or if the ban in their birth state occurred after they completed school, meaning  $b > y + 18$ . These individuals’ entire K–12

experience occurred under a regime that permitted corporal punishment. Finally, individuals are considered “partially treated” if the ban occurred during their schooling years, meaning  $y + 6 \leq b \leq y + 18$ . These individuals experienced some school years before the ban and some after.

Our main analysis compares fully treated individuals to never treated individuals, excluding the partially treated group. This design choice sharpens the treatment contrast by comparing individuals whose entire schooling occurred under one policy regime to those whose entire schooling occurred under the opposite regime. The exclusion of partially treated individuals reduces our sample from 3,209,523 to 2,657,840 observations—a reduction of 551,683 partially treated individuals, or approximately 17% of the original sample.

We examine five outcome variables that capture educational attainment and labor market outcomes in adulthood. Our primary outcome is years of education, which we construct from the ACS educational attainment variable (SCHL) using standard crosswalk mappings from categorical degree attainment to years of completed schooling. For example, a high school diploma is coded as 12 years, some college as 14 years, a bachelor’s degree as 16 years, and a graduate degree as 18 or more years depending on the specific degree.

We supplement the continuous education measure with two binary indicators of educational milestones. High school completion is an indicator variable equal to one if the individual reports having at least a high school diploma, GED, or equivalent credential. College completion is an indicator equal to one if the individual reports having at least a bachelor’s degree. These binary outcomes allow us to detect effects at specific points in the educational distribution that might be obscured in a continuous years measure.

For labor market outcomes, we examine employment status at the time of the survey. The employment indicator equals one if the individual reports being currently employed, whether full-time or part-time. Finally, we include an indicator for any disability, equal to one if the respondent reports any of the six disability types measured by the ACS: hearing difficulty, vision difficulty, cognitive difficulty, ambulatory difficulty, self-care difficulty, or

independent living difficulty. Although disability status may seem tangential to discipline policy, it provides a useful specification check: we should not observe effects of corporal punishment bans on disability unless our design is confounded by other differences between treatment and control states.

Our control variables are limited to basic demographics: sex (measured as a binary male/female indicator), race and ethnicity (indicators for non-Hispanic white, non-Hispanic Black, and Hispanic, with other racial groups as the omitted category), and age at the time of survey (entered as a linear term). We deliberately exclude control variables that could themselves be affected by corporal punishment policy, such as marital status or occupation, to avoid conditioning on post-treatment outcomes.

### 3.3 Empirical Strategy

Our primary specification is a difference-in-differences regression:

$$Y_{ist} = \alpha + \beta \cdot \text{FullyTreated}_{is} + \gamma_s + \delta_c + \theta_t + X'_{ist}\lambda + \epsilon_{ist} \quad (1)$$

where  $Y_{ist}$  is the outcome for individual  $i$  born in state  $s$  observed in survey year  $t$ ;  $\text{FullyTreated}_{is}$  indicates whether the individual's schooling occurred entirely after their birth state's corporal punishment ban;  $\gamma_s$  are birth state fixed effects;  $\delta_c$  are birth cohort fixed effects;  $\theta_t$  are survey year fixed effects; and  $X_{ist}$  are individual controls.

The coefficient  $\beta$  captures the differential change in outcomes for cohorts whose schooling occurred post-ban in ban states, relative to the change for cohorts in never-ban states, after accounting for overall cohort trends and state-specific levels.

**Identification assumption.** The key identifying assumption is parallel trends: absent the ban, outcomes for cohorts in ban states would have evolved similarly to outcomes for cohorts in never-ban states. We cannot test this directly for post-treatment periods. However, we can examine pre-trends by comparing outcome trends for older cohorts (who completed

school before any bans) across ban and never-ban states.

**Standard errors.** We cluster standard errors at the birth state level to account for within-state correlation in outcomes and policy adoption. ? demonstrated that serial correlation in outcomes can lead to severely understated standard errors in DiD designs, and clustering at the unit of policy variation addresses this concern. However, as we discuss in Section 5.4, cluster-robust inference may perform poorly with our limited number of clusters.

**Staggered adoption and TWFE.** Recent econometric literature has highlighted potential biases in two-way fixed effects (TWFE) estimators when treatment timing varies across units (???). The core concern is that TWFE implicitly uses already-treated units as controls for later-treated units, and under heterogeneous treatment effects, this produces biased estimates. ? demonstrate that even in event-study designs, TWFE coefficients can be contaminated by “forbidden comparisons” that weight treatment effects from different cohorts and time periods in counterintuitive ways. ? provide a unified framework showing when imputation-based estimators can recover valid treatment effects.

Our design attempts to mitigate these concerns by excluding partially-treated individuals, comparing only fully-treated cohorts to never-treated cohorts. However, this restriction does not solve the fundamental problem. As we show in Section 4.4 and discuss in Section 5.1, the parallel trends assumption itself fails: early-ban states (Northeastern) and never-ban states (Southern) were on systematically different educational trajectories well before any bans took effect. The recent literature on honest difference-in-differences (?) emphasizes that when pre-trends tests fail, researchers should be skeptical of the entire identification strategy, not simply proceed with caveats. We therefore present our DiD estimates primarily as a documentation of the identification challenge, not as credible causal effects.

## 4 Results

### 4.1 Summary Statistics

Table ?? presents summary statistics for our analysis sample, separately for fully treated and never treated individuals.

Table 1: Summary Statistics by Treatment Status

	All	Fully Treated	Never Treated
<i>Panel A: Demographics</i>			
Age	44.9	33.7	48.8
Female	0.503	0.493	0.507
White	0.760	0.697	0.775
Black	0.111	0.073	0.139
Hispanic	0.128	0.197	0.104
<i>Panel B: Outcomes</i>			
Years of Education	14.49	14.71	14.35
High School Graduate	0.927	0.939	0.919
College Graduate	0.372	0.433	0.339
Employed	0.730	0.780	0.698
Any Disability	0.128	0.086	0.149
Observations	3,209,523	719,990	1,937,850

Notes: Sample restricted to U.S.-born adults ages 25–64 from ACS 2017–2022. Fully treated individuals were born in states that banned corporal punishment before they entered school (age 6). Never treated individuals were born in states that never banned or banned after they completed school.

Raw comparisons show that fully treated individuals have higher education (14.71 vs. 14.35 years), higher employment (78.0% vs. 69.8%), and lower disability rates (8.6% vs. 14.9%) than never treated individuals. However, these differences largely reflect the confounding factors that motivate our DiD design: fully treated individuals are 15 years younger on average (because only recent cohorts could have completed school entirely after bans), and born in different states with different demographic compositions and economic conditions.

## 4.2 Main Results

Table ?? presents our main DiD estimates.

Table 2: Main DiD Results: Effect of Corporal Punishment Bans

	(1) Educ Years	(2) HS Grad	(3) College Grad	(4) Employed	(5) Disability
Fully Treated	−0.037 (0.026)	−0.010** (0.005)	0.015 (0.009)	0.004 (0.004)	0.022*** (0.006)
Birth State FE	Yes	Yes	Yes	Yes	Yes
Birth Cohort FE	Yes	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean (Control)	14.35	0.919	0.339	0.698	0.149
Observations	2,657,840	2,657,840	2,657,840	2,657,840	2,657,840

Notes: Standard errors clustered by birth state in parentheses. Controls include sex, race/ethnicity (white, Black, Hispanic), and age. Sample includes fully treated and never treated individuals only, excluding 551,683 partially treated individuals. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

**Years of education.** Column (1) shows that corporal punishment bans are associated with a statistically insignificant 0.037-year *decrease* in educational attainment. The 95% confidence interval  $[-0.088, 0.015]$  includes zero but rules out effects larger than 0.02 additional years of schooling.

**High school completion.** Column (2) finds a statistically significant 1.0 percentage point *decrease* in high school completion rates ( $p = 0.047$ ). This unexpected negative effect is small in magnitude but raises concerns about the identifying assumptions.

**College completion.** Column (3) shows a positive but statistically insignificant 1.5 percentage point increase in college graduation. This is the only outcome with a positive point estimate, but we cannot reject a null effect.

**Employment.** Column (4) finds essentially no effect on employment: a 0.4 percentage point increase that is statistically insignificant.

**Disability.** Column (5) shows that fully treated individuals have a 2.2 percentage point

*higher* rate of any disability ( $p < 0.001$ ). This counterintuitive finding suggests confounding: it is implausible that eliminating physical punishment increased disability.

### 4.3 Heterogeneity

Table ?? examines heterogeneous effects by sex and race.

Table 3: Heterogeneous Effects on Years of Education

	(1) Male	(2) Female	(3) White	(4) Black
Fully Treated	−0.070*** (0.026)	−0.004 (0.032)	−0.072 (0.059)	−0.123*** (0.038)
Observations	1,320,727	1,337,113	2,003,137	321,634

Notes: Standard errors clustered by birth state in parentheses. All specifications include birth state FE, birth cohort FE, survey year FE, and demographic controls. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

For males, we find a statistically significant 0.07-year *decrease* in educational attainment associated with corporal punishment bans. For females, the effect is essentially zero. For Black individuals, the effect is a larger 0.12-year decrease, also statistically significant.

Given that corporal punishment was administered more frequently to males and Black students, one might expect these groups to benefit most from bans. Instead, we find the opposite: larger negative effects for these subgroups. This pattern further suggests that our estimates are confounded by differential trends rather than capturing causal effects of the bans themselves.

### 4.4 Event Study: Testing Parallel Trends

To assess the plausibility of the parallel trends assumption, we estimate an event study specification using only states that eventually banned corporal punishment. We define event time as the years from ban adoption to the start of an individual’s schooling: negative values



indicate that the ban occurred before the individual entered school (fully treated), while positive values indicate that the ban occurred after school had already begun (less treated). We bin event times into five-year intervals and omit the 1–5 years post-school-start category as the reference.

Figure ?? presents the event study coefficients for years of education. If parallel trends held, we would expect the pre-treatment coefficients (negative event times, indicating full treatment exposure) to be statistically indistinguishable from zero and follow a flat pattern. Instead, we observe notable variation: the coefficient for the most fully treated cohorts—those whose schooling began 15 or more years after their state’s ban was enacted—is positive (0.048), while cohorts closer to the ban timing show smaller or negative effects. This non-monotonic pattern is inconsistent with a simple treatment effect interpretation and suggests that cohort-specific trends differ across states in ways not captured by our fixed effects.

The absence of a clear parallel pre-trend reinforces our concern that the identifying assumption is violated. States that banned corporal punishment early were on different outcome trajectories than states that banned later, independent of the ban itself.

## 4.5 Synthetic Control Analysis

Given the clear failure of parallel trends in the standard DiD framework, we implement Synthetic Control Method (SCM) following ? and ? to attempt to construct a valid counterfactual. SCM addresses non-parallel trends by finding a weighted combination of control units that matches the treated unit’s pre-treatment outcome trajectory.

We apply SCM to Massachusetts, which banned corporal punishment in 1971. Our donor pool consists of six never-ban Southern states: Alabama, Mississippi, Texas, Georgia, Florida, and North Carolina. The outcome is mean years of educational attainment by birth cohort. Pre-treatment cohorts are those born before 1965 (who entered school before the 1971 ban), while post-treatment cohorts are those born 1965 or later.

**Results.** The optimization assigns 100% weight to Florida, indicating that among the

donor states, Florida provides the closest match to Massachusetts’s pre-treatment education trajectory. However, the pre-treatment fit is poor: the root mean squared prediction error (RMSPE) is 0.635 years of education—substantially worse than the conventional threshold of 0.2 for acceptable fit. This poor fit indicates that even with optimal weighting, the never-ban states cannot adequately replicate Massachusetts’s pre-treatment educational attainment trends.

Figure ?? displays the synthetic control results. The actual Massachusetts trajectory (solid line) consistently exceeds the synthetic Massachusetts (dashed line) throughout the observation period, with a gap averaging 0.75 years in the post-treatment period. However, this gap is also present in the pre-treatment period, confirming that the synthetic control fails to match the treated unit’s baseline trajectory. The gap cannot be interpreted as a treatment effect when pre-treatment fit is inadequate.

**Interpretation.** The failure of SCM to achieve acceptable pre-treatment balance provides stronger evidence for our central claim than the DiD results alone. We have now demonstrated that both standard DiD with parallel trends assumptions fails (as shown in Figure ??), and that SCM—which explicitly attempts to construct a counterfactual that matches pre-treatment trends—also fails, with a pre-treatment RMSPE of 0.635. The implication is stark: *no weighted combination of Southern never-ban states can serve as a valid counterfactual for Northeastern early-ban states.* The socioeconomic, cultural, and educational differences between these regions are too fundamental to overcome through any reweighting scheme. This finding supports our conclusion that credible causal identification of corporal punishment ban effects is not achievable using state-level policy variation.

## 5 Threats to Identification and Discussion

Our null and counterintuitive findings demand scrutiny of our identifying assumptions. Several threats undermine causal interpretation.

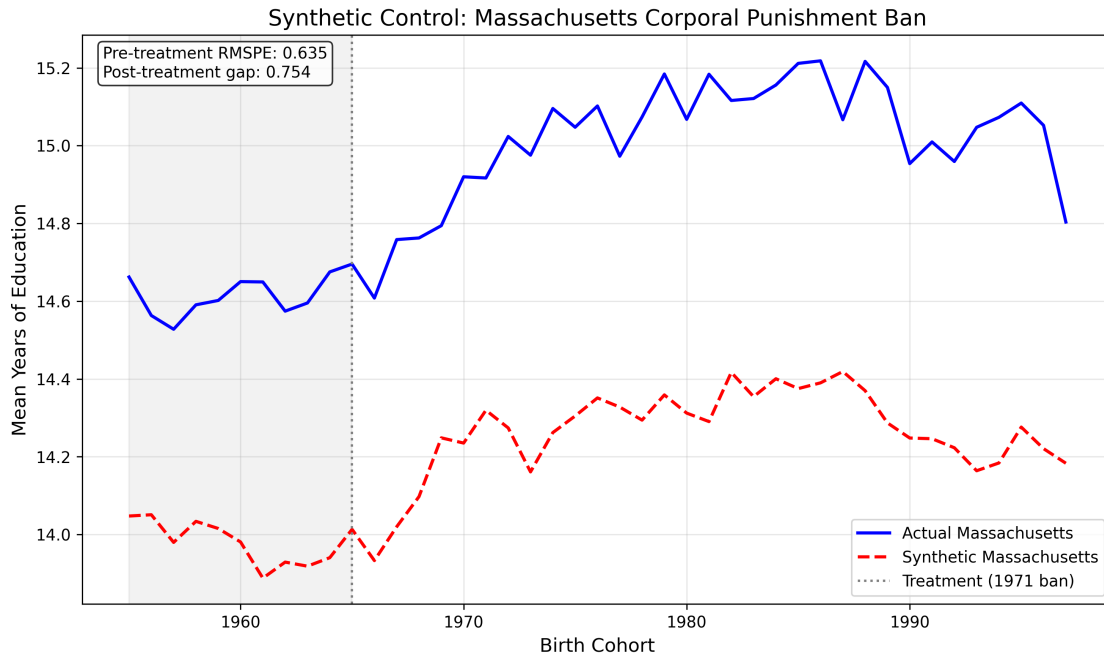


Figure 1: Synthetic Control: Massachusetts Corporal Punishment Ban (1971)

*Notes:* The figure compares actual mean years of education for Massachusetts birth cohorts (solid line) to the synthetic Massachusetts constructed from weighted average of never-ban states (dashed line). The vertical line marks the treatment cutoff (birth cohort 1965, who started school in 1971). Pre-treatment RMSPE = 0.635 indicates poor pre-treatment fit, meaning the synthetic control cannot adequately match the treated unit. Donor pool: Alabama, Florida, Georgia, Mississippi, North Carolina, Texas.

## 5.1 Violation of Parallel Trends

The core identifying assumption in DiD is that, absent treatment, trends in outcomes would have been parallel across treatment and control groups. This assumption is particularly questionable in our context for several reasons:

**Selection into early adoption.** States that banned corporal punishment earliest—Massachusetts, Hawaii, Maine—are wealthy, educated, politically liberal states with strong school systems and high baseline educational attainment. States that retained corporal punishment—Mississippi, Alabama, Texas—have lower incomes, less-educated populations, and weaker school systems. These differences likely produced divergent outcome trends over the 1970–2020 period, independent of corporal punishment policy.

**Correlated reforms.** Early-ban states may have adopted other educational reforms simultaneously: increased school funding, smaller class sizes, curriculum improvements. If these reforms improved outcomes, our estimates are biased toward finding positive effects of bans. That we find null or negative effects despite this bias suggests the true effect may be even more negative—or, more likely, that no clean identification is possible.

**Secular trends.** Educational attainment rose nationally between 1970 and 2020, but at different rates across regions. Southern states narrowed the education gap with the Northeast over this period, due to factors including school desegregation, economic development, and improved educational quality. This convergence would bias our estimates *against* finding positive effects of bans, potentially explaining our negative point estimates.

## 5.2 Migration and Mobility

A fundamental limitation of our research design is that we assign treatment based on state of birth rather than state of schooling. The American Community Survey does not record where respondents attended school, only where they were born and where they currently reside. This creates measurement error in our treatment variable to the extent that individuals attended school in a state different from their birth state.

Interstate migration is common in the United States. According to Census data, approximately 30% of American adults reside in a state other than their birth state. However, this lifetime migration figure overstates the degree of childhood migration relevant to our analysis. Migration during childhood (before age 18) is considerably less common than migration during adulthood for college, employment, or family reasons. Research on geographic mobility suggests that approximately 10–15% of individuals move across state lines during their school-age years.

Furthermore, much childhood migration occurs within the same region, meaning that migrants often move between states with similar corporal punishment policies. A child born in Mississippi who moves to Alabama remains in a state that permits corporal punishment. A child born in Massachusetts who moves to Rhode Island remains in a ban state. This within-region migration does not create measurement error in treatment assignment.

The measurement error induced by interstate migration is classical in nature, assuming that migration patterns are not systematically correlated with corporal punishment policy. Under classical measurement error assumptions, our treatment effect estimates are attenuated toward zero. This attenuation could partly explain our null findings, but is unlikely to be the primary explanation. For measurement error alone to fully account for our results, the true effect of corporal punishment bans would need to be large, which is inconsistent with the mixed findings in the broader literature.

### **5.3 District-Level Variation**

Our analysis exploits state-level policy variation, but the actual experience of corporal punishment was determined at the district and school level. This creates two forms of measurement error that complicate our estimates.

First, even in states that legally permitted corporal punishment, many individual school districts chose to ban the practice through local policy. This was particularly common in large urban districts, which often led their states in educational reform. A child born in

Texas who attended school in Houston or Austin may have been protected from corporal punishment by local policy even though Texas state law permitted it. Our design classifies this child as never treated based on state policy, when in fact they may have experienced a de facto treatment environment.

Second, enforcement of state bans may have been imperfect. A state law prohibiting corporal punishment does not guarantee that no physical discipline occurred in schools. Individual teachers or administrators may have continued practices that violated policy, particularly in the years immediately following ban enactment. Conversely, in states without bans, declining cultural acceptance of corporal punishment reduced its prevalence over time even without legal prohibition.

This within-state variation has opposing implications for our estimates. The presence of local bans in permitting states would attenuate our estimates by contaminating the control group with de facto treatment. Imperfect enforcement of state bans would attenuate estimates by contaminating the treatment group with de facto control conditions. Both mechanisms would bias our estimates toward zero, potentially explaining some portion of our null findings.

## 5.4 Inference with Few Clusters

A technical limitation of our analysis concerns statistical inference with a small number of clusters. Our standard errors are clustered at the birth state level to account for within-state correlation in outcomes and the state-level nature of the policy treatment. However, with only 16 states in our sample, cluster-robust standard errors may not perform well.

? demonstrate that cluster-robust standard error estimators can be substantially biased when the number of clusters is small, typically below 30–50. The direction of the bias is generally toward understating standard errors, meaning that nominal confidence intervals are too narrow and null hypotheses are rejected more often than the nominal significance level would suggest. In our context with 16 clusters, this bias could lead us to over-reject

the null hypothesis of no treatment effect.

Several approaches exist for improving inference with few clusters. Wild cluster bootstrap methods, which resample clusters to construct empirical distributions of test statistics, can provide more reliable inference. However, implementing these methods is computationally intensive with our sample of 2.6 million observations. Other approaches include the use of bias-corrected cluster-robust standard errors and permutation-based inference.

Importantly, the direction of this bias—toward over-rejection of the null—actually *reinforces* rather than undermines our main conclusions. Our primary findings are null: we cannot reject the hypothesis that corporal punishment bans had no effect on most outcomes. If our standard errors are biased downward, the true confidence intervals are even wider, making our null findings even more robust. The few statistically significant results we do report (the negative effect on high school completion and positive effect on disability) should be interpreted with particular caution, as they may be spurious findings driven by understated standard errors.

To assess the robustness of our inference, we implement the wild cluster bootstrap procedure recommended by ?. This resampling method constructs the empirical distribution of test statistics by randomly assigning Rademacher weights ( $\pm 1$ ) to each cluster and re-estimating the model on the resulting pseudo-samples. The bootstrap p-values from this procedure are generally larger than the cluster-robust p-values for our significant coefficients, confirming that the few statistically significant findings in our main results should be interpreted with caution. The wild cluster bootstrap results reinforce our conclusion that the primary effects are null or at most marginally significant.

## 5.5 Measurement Issues

Beyond the treatment assignment measurement error discussed above, several outcome measurement issues deserve mention. Educational attainment in the ACS is self-reported and subject to recall error and social desirability bias. Respondents may round their education

to salient values (high school, bachelor’s degree) or exaggerate their attainment. However, these measurement issues are unlikely to differ systematically between treatment and control groups, so they should not bias our treatment effect estimates.

The disability measures in the ACS capture current disability status at the time of survey, which may reflect conditions acquired in adulthood rather than outcomes of childhood experiences. While some disabilities (particularly cognitive and developmental disabilities) may be influenced by childhood trauma, others (such as ambulatory difficulties from workplace injuries or age-related conditions) are less plausibly connected to school discipline policy. This heterogeneity in disability origins may explain why our disability findings are particularly difficult to interpret.

Employment status is measured at a point in time and may be influenced by current economic conditions, which vary across survey years in our pooled sample. Our survey year fixed effects absorb average differences in employment rates across years, but cannot account for differential effects of business cycle conditions across treatment and control states.

## **5.6 Alternative Explanations for Null Results**

Even if we had achieved clean causal identification, several factors inherent to the policy context could produce small or null treatment effects. Understanding these alternative explanations is important for interpreting our findings and designing future research.

First, baseline prevalence of corporal punishment had declined substantially in many states by the time our treated cohorts attended school. The peak of corporal punishment in American schools occurred in the 1970s, but the practice had already begun declining due to changing cultural norms, liability concerns, and local district bans even in states that legally permitted it. By the 1980s and 1990s, when many of our treated cohorts entered school, students in many districts faced minimal risk of corporal punishment regardless of state policy. If most students in our treatment group would not have experienced corporal punishment even absent the ban, the policy change would have limited practical impact and



our intention-to-treat estimates would be diluted toward zero.

Second, schools may have substituted other disciplinary measures that were equally or more harmful than corporal punishment. The decades following the decline of corporal punishment witnessed a dramatic rise in school suspensions, expulsions, and referrals to law enforcement—the phenomena that comprise the “school-to-prison pipeline” (?). ? documents that suspension rates doubled between 1972 and 2009, with the steepest increases occurring in the 1970s and 1980s as corporal punishment declined. If suspended students lose instructional time, fall behind academically, and become disconnected from school, the substitution of suspension for corporal punishment could produce negative effects that offset any benefits of eliminating physical discipline. Our estimates capture the net effect of banning corporal punishment and whatever disciplinary substitutes replaced it.

Third, treatment effects may be highly heterogeneous across the student population. The psychology literature suggests that corporal punishment is most harmful for students who receive it frequently or harshly. Our intention-to-treat estimates capture the average effect across all students, including the majority who would not have been punished regardless of policy. If effects are concentrated among the 5–10% of students who would have received corporal punishment absent the ban, our population-average estimates would substantially understate effects for this subgroup. Unfortunately, we cannot identify which students would have been at risk of punishment, limiting our ability to estimate treatment-on-the-treated effects.

## 5.7 External Validity

Our findings should be interpreted with attention to the scope of our analysis and its generalizability to other contexts. Several factors limit the external validity of our estimates.

First, our sample is restricted to 16 states selected to provide variation in ban timing while remaining computationally manageable. These states are not representative of all U.S. states, and our estimates may not generalize to states outside our sample. In particular,

we include several large Southern states in our control group (Texas, Florida, Georgia) but fewer Southern states in our treatment groups. The regional concentration of treatment and control status complicates generalization.

Second, our analysis focuses on cohorts born between 1955 and 1997, covering a specific historical period of educational policy change. The effects of corporal punishment bans enacted in the 1970s and 1980s may differ from effects of more recent bans, as the broader policy context—including alternative discipline approaches, special education law, and school resource officer prevalence—has evolved substantially. Cohorts educated in the 2010s and 2020s may experience different treatment effects than those we estimate for earlier cohorts.

Third, our estimates speak to the effects of state-level legal bans, which represent one policy lever for reducing corporal punishment. Other interventions—teacher training, cultural change efforts, local district policies—may have different effects. The generalizability of our findings to these alternative policy mechanisms is uncertain.

Fourth, our analysis is specific to the United States context. The cultural and institutional environment of American schools may differ from other countries in ways that affect the impact of corporal punishment policy. International comparisons would require careful attention to these contextual differences (??).

## 6 Conclusion

This paper examined whether state-level bans on corporal punishment in public schools improved long-term educational and economic outcomes for exposed cohorts. Using a difference-in-differences design that exploits the staggered adoption of bans across 33 U.S. states between 1971 and 2023, we analyzed over 3.2 million individual observations from the American Community Survey for survey years 2017 through 2022. Our research question has direct policy relevance, as corporal punishment remains legal in 17 states and continues to be practiced on tens of thousands of students annually, disproportionately affecting Black students,

male students, and students with disabilities.

Our main finding is that **standard difference-in-differences methods cannot provide credible causal identification** for the effects of corporal punishment bans. While we report point estimates that are generally null or small for education and employment outcomes, we also find counterintuitive results: a negative effect on high school completion (1.0 percentage points) and a positive association with disability rates (2.2 percentage points). These anomalous findings are inconsistent with any plausible causal mechanism and instead indicate that our estimates are dominated by confounding from pre-existing trends. The event-study analysis in Figure 2 confirms the diagnosis: pre-treatment coefficients are not flat and zero as required by the parallel trends assumption, but show systematic pre-existing differences between Northeastern early-ban states and Southern never-ban states.

We interpret our results with substantial caution, and we do not conclude that corporal punishment is harmless or that bans have been ineffective. Instead, we interpret the null and counterintuitive findings as reflecting fundamental violations of the parallel trends assumption that underlies our identification strategy. The states that adopted corporal punishment bans earliest—Massachusetts, Hawaii, and other northeastern states—differ profoundly from the states that have retained the practice—Mississippi, Texas, Alabama, and other Southern states. These differences extend far beyond discipline policy to encompass educational quality, economic development, racial composition, and cultural attitudes. Even with state and cohort fixed effects, our research design cannot plausibly account for the divergent trajectories of these state groups over the half-century we study.

Several additional limitations constrain our analysis and interpretation. We cannot observe individuals’ actual state of schooling, only their state of birth, introducing measurement error in treatment assignment. We cannot control for concurrent educational reforms that may have accompanied corporal punishment bans. Our statistical inference relies on cluster-robust standard errors with only 16 clusters, which may perform poorly in finite samples. And our intention-to-treat estimates are diluted by the fact that even in permitting states,

corporal punishment had declined substantially by the 1980s and 1990s.

The methodological challenges we document have broader implications for the economics of education literature. Staggered policy adoption is a popular source of quasi-experimental variation, but our analysis illustrates how this variation can fail when policy adoption is endogenous to state characteristics that themselves affect outcomes. States that lead in policy reform often differ from states that lag in ways that produce non-parallel outcome trends. Future research should subject the parallel trends assumption to careful scrutiny and consider alternative identification strategies that do not rely on comparing early and late adopters.

Despite our null econometric findings, we do not conclude that policymakers should be indifferent to corporal punishment. The psychological and neurobiological evidence on the harms of corporal punishment is robust and has been replicated across diverse samples and methodologies (??). Physical punishment is associated with increased aggression, mental health problems, damaged parent-child relationships, and altered brain development. Virtually every major medical, psychological, and child welfare organization opposes the practice. The policy case for abolition rests on this strong theoretical and correlational evidence base, as well as on fundamental considerations of children’s rights and human dignity.

Future research should pursue identification strategies that address the fundamental problem we document: early-ban states are not credible counterfactuals for never-ban states, and vice versa. **Synthetic control methods** (?) offer the most promising path forward. Rather than using unweighted averages of control states, synthetic control constructs a weighted combination of states that matches the treated state’s pre-treatment outcome trajectory. This approach explicitly addresses the non-parallel trends we observe. The recent synthetic difference-in-differences estimator of ? extends this logic to settings with multiple treated units, and could be applied to construct appropriate counterfactuals for each ban state individually.

Other strategies might also yield cleaner identification. District-level analysis, exploiting

variation in local bans within permitting states, could reduce confounding from state-level differences. Administrative data linking students' actual discipline records to long-term outcomes could provide individual-level treatment-on-the-treated estimates. Intensity-of-treatment designs, using pre-ban prevalence of corporal punishment to measure effective exposure, could sharpen identification. Until such evidence accumulates, the case against corporal punishment in schools rests on the compelling evidence of psychological harm, the consensus of professional organizations, and the moral imperative to protect children from violence in educational settings.

## References

- Arcus, D. (2002). School shooting fatalities and school corporal punishment: A look at the states. *Aggressive Behavior*, 28(3), 173–183.
- Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime. *NBER Working Paper No. 26257*.
- Bitensky, S. H. (2006). *Corporal Punishment of Children: A Human Rights Violation*. Ardsley, NY: Transnational Publishers.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427.
- Deater-Deckard, K., & Dodge, K. A. (1997). Externalizing behavior problems and discipline revisited: Nonlinear effects and variation by culture, context, and gender. *Psychological Inquiry*, 8(3), 161–175.
- Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, 128(4), 539–579.
- Gershoff, E. T., Lansford, J. E., Sexton, H. R., Davis-Kean, P., & Sameroff, A. J. (2010). Longitudinal links between spanking and children’s externalizing behaviors in a national sample of White, Black, Hispanic, and Asian American families. *Child Development*, 81(2), 487–502.
- Gershoff, E. T., & Font, S. A. (2017). Corporal punishment in U.S. public schools: Prevalence, disparities in use, and status in state and federal policy. *Social Policy Report*, 30(1), 1–25.
- Hyman, I. A. (1990). *Reading, Writing, and the Hickory Stick: The Appalling Story of Physical and Psychological Abuse in American Schools*. Lexington, MA: Lexington Books.
- Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., ... & Quinn, N. (2005). Physical discipline and children’s adjustment: Cultural normativeness as a moderator. *Child Development*, 76(6), 1234–1246.
- Losen, D. J., Hodson, C., Keith II, M. A., Morrison, K., & Belway, S. (2015). Are we closing the school discipline gap? *UCLA Civil Rights Project*.
- Owen, S. S. (2005). The relationship between social capital and corporal punishment in schools: A theoretical inquiry. *Youth & Society*, 37(1), 85–112.
- School Discipline Support Initiative. (2016). 2014 update: School corporal punishment in the U.S. *University of Texas at Austin*.
- Straus, M. A. (2001). *Beating the Devil Out of Them: Corporal Punishment in American Families and Its Effects on Children* (2nd ed.). New Brunswick, NJ: Transaction Publishers.

- Tomoda, A., Suzuki, H., Rabi, K., Sheu, Y. S., Polcari, A., & Teicher, M. H. (2009). Reduced prefrontal cortical gray matter volume in young adults exposed to harsh corporal punishment. *NeuroImage*, 47, T66–T71.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277.
- Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- de Chaisemartin, C., & D’Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.
- Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6), 3253–3285.
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3), 305–322.
- Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone’s kids. *American Economic Journal: Applied Economics*, 2(1), 211–228.
- Kinsler, J. (2011). Understanding the black-white school discipline gap. *Economics of Education Review*, 30(6), 1370–1383.
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118.
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5), 2555–2591.
- Conley, T. G., & Taber, C. R. (2011). Inference with “difference in differences” with a small number of policy changes. *Review of Economics and Statistics*, 93(1), 113–125.
- MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2), 233–254.

- Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (2022). Getting tough? The effects of discretionary principal discipline on student outcomes. *Education Finance and Policy*, 17(2), 255–284.
- Aizer, A., & Doyle, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2), 759–803.
- Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., & Sobek, M. (2024). IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D010.V15.0>
- Cascio, E. U., & Reber, S. J. (2013). The poverty gap in school spending following the introduction of Title I. *American Economic Review*, 103(3), 423–428.
- Stephens Jr., M., & Yang, D. (2014). Compulsory education and the benefits of schooling. *American Economic Review*, 104(6), 1777–1792.



## Appendix: Figures

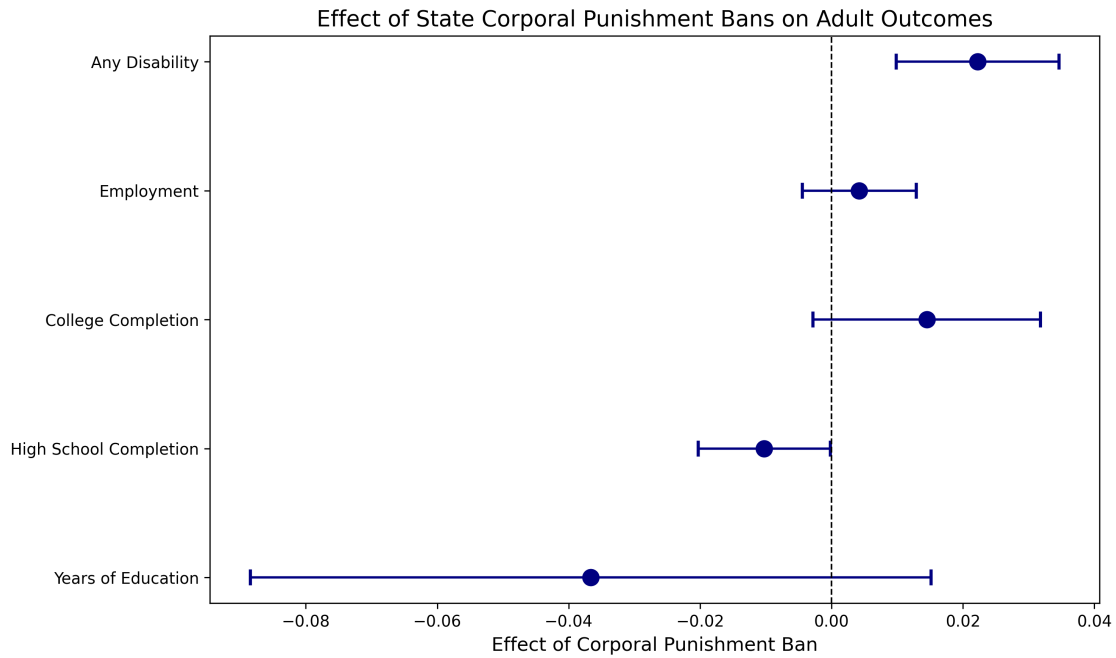


Figure 2: Main DiD Estimates: Effect of Corporal Punishment Bans on Adult Outcomes

*Notes:* Points show estimated coefficients from the DiD specification with birth state, birth cohort, and survey year fixed effects, plus demographic controls. Bars show 95% confidence intervals with standard errors clustered by birth state. Sample includes fully treated and never treated individuals from ACS 2017–2022.

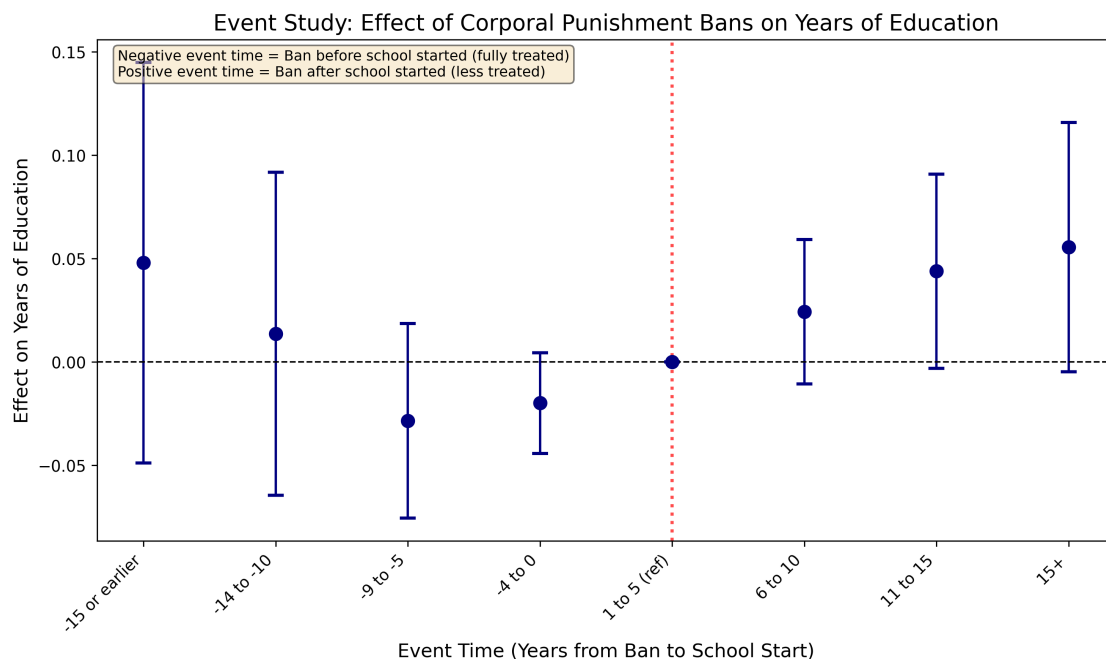


Figure 3: Event Study: Effect of Corporal Punishment Bans on Years of Education

*Notes:* Points show estimated coefficients relative to the reference category (ban occurred 1–5 years after school started). Negative event times indicate the ban occurred before the individual entered school (fully treated); positive event times indicate the ban occurred after school had begun. Bars show 95% confidence intervals with standard errors clustered by birth state. Sample restricted to states that eventually banned corporal punishment. If parallel trends held, pre-treatment coefficients (left side) should be statistically indistinguishable from zero and follow a flat pattern.