

Community Impacts of Mass Incarceration*

Arpit Gupta[†]

NYU Stern

Christopher Hansman[‡]

Imperial College London

Evan Riehl[§]

Cornell University

May 9, 2023

Abstract

We show community-level incarceration rates negatively impact the academic performance of local students, including children without direct exposure to an incarcerated household member. Our identification strategy exploits the year-to-year turnover of judges who vary in their tendency to incarcerate, and uses a new dataset that matches students' test scores to defendants' court records at the address level. We highlight one mechanism for this effect: spillovers from directly exposed children onto a broad set of peers within the classroom. The results underscore important consequences of incarceration for access to opportunity within entire communities.

JEL codes: K14, K42, I24, J13

*We thank Jason Baron, Steve Billings (discussant), Kara Bonneau, Jamein Cunningham, Ryan Fackler, Ethan Frenchman, David Hoke, Naci Mocan (discussant), Bentley MacLeod, Michael Mueller-Smith, Aurélie Ouss (discussant), John Rubin, Jessica Smith, and Megan Stevenson for helpful conversations and conference participants at SOLE, Texas Economics of Crime Workshop, NYU Law, NBER Law & Economics, ALEA, TWEC, and UIUC. We also thank Dalya Elmalt, Joshua Coven, Abhinav Gupta, and Harry Kosowsky for superb research assistance. We are grateful to the North Carolina Education Research Data Center for its data assistance. This research received financial support from the Alfred P. Sloan Foundation through the NBER Household Finance small grant program. The protocol for this study has been approved by the NYU IRB (#1242).

[†]Department of Finance, NYU Stern School of Business; Email: arpit.gupta@stern.nyu.edu; Web: arpitgupta.info.

[‡]Department of Finance, Imperial College Business School, Imperial College London; Email: chansman@imperial.ac.uk; Tel: +44 (0)73 80320232; Web: <https://chansman.github.io/>.

[§]Department of Economics and ILR School, Cornell University; Email: eriehl@cornell.edu; Tel: +1 (607) 255-0395; Web: <http://riehl.economics.cornell.edu/>.

1 Introduction

The unprecedented growth in incarceration since the 1970s has left more than two million adults in jails or prisons in the United States, and underprivileged communities have borne the brunt of this expansion. Mass incarceration disrupts local social norms and networks (Clear, 2008), which may partly explain why heavily exposed areas have lower intergenerational mobility (Chetty, Hendren, Kline, and Saez, 2014; Chetty, Friedman, Hendren, Jones, and Porter, 2018). However, establishing causal relationships at the community level is challenging because it is difficult to credibly separate incarceration rates from criminal activity, household sorting, and other neighborhood-level factors. As a result, prior work in economics has primarily focused on how incarceration affects criminal defendants and their family members, with less attention paid to impacts on the wider community.

In this paper we investigate the impact of incarceration rates on a key community-level outcome: academic achievement for local students. To do so, we use a novel dataset that links criminal justice and education records for the entire state of North Carolina by matching defendant and student addresses. Our central empirical strategy is a judge turnover design that exploits year-to-year changes in the composition of judges in each county court. Specifically, we focus on the entry and exit of judges who differ in stringency (their tendency to incarcerate criminal defendants).¹ Turnover has a significant impact on local incarceration—a one standard deviation increase in county-level stringency leads to a 15–20 percent increase in incarceration—but is not related to trends in criminal activity, defendant demographics, or local students’ characteristics.

Our main finding is that increases in local incarceration lead to meaningful reductions in student achievement. A one standard deviation increase in county-level stringency reduces average math scores across the community by 2.5–3.5 percent of a standard deviation (and reduces English scores by 1.5–2.0 percent of a standard deviation). These effects are persistent, growing in magnitude over the course of at least three years. Crucially, they are sizeable even when considering students whose households have no direct exposure to the criminal justice system. This suggests

¹ Judges enter and exit as a consequence of elections, retirements, promotions, and cross-county rotations. Our judge turnover design is similar in spirit to the empirical strategy in Chetty, Friedman, and Rockoff (2014) that exploits the arrival and departure of teachers who vary in value added. It builds on work using variation in the stringency of judges assigned to individual defendants (e.g., Mueller-Smith, 2015) but focuses on aggregate variation within counties over time.

that the aggregate consequences of incarceration must be driven, at least in part, by spillovers onto a broad set of children in the community.

The remainder of the paper investigates the mechanisms driving this community-wide effect. We provide evidence that a portion of the impact operates through a classroom disruption channel: spillovers from students who directly experience an incarceration in the household onto their classmates. To do so, we examine the impacts of incarceration on directly-affected children using two empirical strategies that are common in the literature. The first is an event study that compares a student's outcomes before and after a household incarceration, relative to other students with household members convicted of the same crime but not incarcerated. The second is a within-court judicial randomization approach that isolates judge-driven variation in incarceration for criminal defendants charged with equivalent crimes. Tests for pre-trends and balance on observables support the identifying assumptions for both approaches.

Both approaches indicate that an incarceration in the household increases classroom misbehavior. For instance, our event studies reveal that affected students experience a 15 percent increase in suspension days, an additional half day absent per year, and a greater number of reported fighting incidents. We also find sizable impacts on academic outcomes—a household incarceration reduces students' math and English scores by 1.0–1.5 percent of a standard deviation.² As the literature studying classroom-level externalities generated by troubled children (e.g., [Lazear, 2001](#); [Carrell and Hoekstra, 2010](#)) points out, this sort of behavioral disruption may substantially impact the academic achievement of classmates.

Indeed, while our sample is not designed to precisely measure relationships between students, our last step provides suggestive evidence in favor of a behavioral disruption channel. We repeat our event study and judge randomization strategies but consider outcomes for students who share a school and grade level with a directly affected child (and are not directly impacted themselves). The event studies suggest that this form of indirect exposure reduces math and English scores by 0.4 and 0.3 percent of a standard deviation, respectively. While these spillover estimates are relatively modest per student, a single child experiencing a household incarceration will interact with a large number of classmates. As a result, classroom spillovers aggregate to explain a meaning-

²These academic declines alone cannot explain the aggregate test-score reductions found in our judge-turnover approach. While directly impacted students do appear to suffer academically, their proportion in the population is far too small to meaningfully impact average academic outcomes given the order of magnitude of the estimated effects.

ful fraction of the relationship between incarceration and student achievement at the community level.

Back of the envelope calculations suggest that the classroom disruption channel alone can explain 6–9 percent of the unconditional aggregate relationship between incarceration and test scores (and roughly 15 percent conditional on observables). A significant portion of the correlation is still unaccounted for, but this is unsurprising. Classmates make up only a fraction of a student’s relationships—they interact with students in other grades and with both children and adults in a variety of non-academic settings—so we expect that the classroom channel represents just one of many mechanisms through which incarceration causally impacts community level outcomes. Taken collectively, our results suggest that impacts on academic achievement extend well beyond the traditional boundaries of the household.

Our findings do not provide a full picture of the costs and benefits of incarceration, but they highlight an important consideration in this discussion. We focus on a particular set of student outcomes and on the specific context of the criminal justice system in North Carolina. It is possible that alternative approaches to rehabilitation and reintegration could lead to more positive outcomes for criminal defendants, their family members, and communities. Additionally, criminal sentencing may play an important role in limiting criminal activity through deterrence and incapacitation. Instead, our results highlight the widespread consequences of high incarceration rates for local student achievement. Policymakers must consider broader spillovers within communities when weighing the total costs and benefits of incarceration.

Our paper contributes to several literatures but relates most closely to interdisciplinary research on the relationship between incarceration and a variety of community outcomes. This work has emphasized the disproportionate local impacts that result from the spatially concentrated nature of incarceration, with lasting consequences along multiple dimensions. This includes the relationship between parental incarceration and children’s behavioral outcomes, such as aggression, which could affect the achievement of peers or classmates.³ Incarceration may also ripple out into local areas by disrupting social networks, affecting marriage markets (Thomas and Sawhill, 2005), and altering family relations during and after the sentence itself (Pattillo, Western, and Weiman,

³ See Wildeman (2010); Geller, Cooper, Garfinkel, Schwartz-Soicher, and Mincy (2012); Wildeman and Turney (2014); Haskins (2014); Hagan and Dinovitzer (1999); Murray and Farrington (2005); Foster and Hagan (2007).

2004). Other community impacts of incarceration include those on labor markets (Larson, Shannon, Sojourner, and Uggen, 2021), local household income (Murray, 2013), and broader shifts in social norms toward family formation, authorities, and the government (Rose and Clear, 1998; Lynch and Sabol, 2004). Finally, the cumulative consequences of incarceration may further generate local criminal activity in the long run (Clear, Rose, Waring, and Scully, 2003). We draw on this large literature and contribute to it by providing causal evidence on the existence of community-level impacts and by highlighting a classroom disruption channel.

A further contribution of this paper is separating the spillover effects of incarceration from other factors that affect children's academic outcomes. Research has found that children's achievement is lower when they have classmates who have experienced domestic violence (Carrell and Hoekstra, 2010; Carrell, Hoekstra, and Kuka, 2018) or the arrest of a parent (Billings and Hoekstra, 2019). Finlay, Mueller-Smith, and Street (2023) show that children experience negative outcomes when their families move to neighborhoods with high rates of exposure to the criminal justice system. Criminal activity, arrests, family demographics, and other dimensions of school and neighborhood quality are each strongly correlated with local incarceration rates, making it difficult to isolate the role of sentencing decisions. While these other factors are surely important, our findings suggest there are widespread consequences of incarceration itself across communities.

We also contribute to an economics literature focusing on the direct impact of parental incarceration on children, which has previously found mixed results. Several papers use event study designs that ask whether children's outcomes change after a parent's incarceration. Cho (2009a,b) and Billings (2018) find either no change or modest improvements in educational and behavioral outcomes. Other work uses judge randomization designs to estimate the effects of parental incarceration, primarily in contexts outside the United States. For example, Dobbie, Grönqvist, Niknami, Palme, and Priks (2018) find negative impacts on the grades and longer-term outcomes of Swedish children, while Bhuller, Dahl, Loken, and Mogstad (2018) find insignificant effects on grades in Norwegian data. Arteaga (2021) finds positive effects of incarceration on educational attainment in Colombia. Perhaps closest to our paper, Norris, Pecenco, and Weaver (2021) find decreases in later life incarceration and improvements in neighborhood quality for directly impacted students in Ohio.

A multidimensional view of the effects of incarceration can help reconcile our results on di-

rectly impacted students with this literature. As [Murray and Farrington \(2008\)](#) emphasize, childhood exposure to parental incarceration, while potentially harmful, may also lower children's crime rates by removing negative role models or via deterrence. A child can simultaneously be traumatized by the incarceration of a parent and be steered away from a criminal path. Indeed, for completeness, we follow [Norris et al. \(2021\)](#) in considering crime rates for children themselves and find consistent estimates, although our sample permits a more powerful test when focusing on behavioral and academic impacts.

More fundamentally, however, our central contribution to this literature is to shift focus toward the broader consequences of incarceration for entire communities. While directly affected children may be most acutely impacted, a substantially larger fraction of students are exposed to the criminal justice system through indirect channels. Our county-level approach provides novel evidence on the aggregate impacts of incarceration for student achievement. The resulting estimates are an order of magnitude too high to be explained through a direct channel alone, but even modest spillover effects inside and outside the classroom can aggregate to explain a sizable fraction of the community achievement-incarceration gradient.

2 Data and Background: the Achievement-Incarceration Gradient

Estimating the impacts of incarceration on community-level educational outcomes is a challenging problem requiring detailed and comprehensive data that link criminal defendants to students. We address these hurdles using a new merge between administrative datasets in North Carolina. We focus on North Carolina because it provides extensive administrative data on courts and public schools and because it is broadly representative of communities around the country in terms of demographic composition and exposure to incarceration.

By construction, our merge connects criminal defendants to children based on address. This enables us to focus on the role of incarceration at the household level, whereas prior literature has typically linked criminal defendants to children by birth records. The incarceration of a household member (versus a parent) may be more disruptive, on average, since many children do not live with both of their parents. For example, in 2010, roughly one in four U.S. children—and more

than half of Black children—lived without a father in the household.⁴

2.1 Data Sources

We use administrative criminal justice and education data from two agencies in North Carolina.

Court Records (ACIS). The first data source is the North Carolina Court’s Automated Criminal/Infractions System (ACIS). These records cover all criminal cases in the state for which the date of last update was between July 1, 2009 and June 30, 2014. The data allow us to track the progress of individuals who interact with the criminal justice system from arrest to sentencing, and provides defendant demographics and characteristics of the case at origination such as the date, county, court type (district or superior), and all offense codes at arrest. Importantly for our strategy, the data additionally include the initials of the judge who handled the case and the date of disposition.

We also observe extensive information on sentencing outcomes, including the type of sentence, offense codes and structured sentencing offense class at conviction, and the defendant’s prior points. Throughout the paper, we exclude lower-level traffic offenses (classes 2–3), which rarely result in incarceration. Appendix C provides a detailed overview of our data. Appendix C.1 provides variable definitions, Appendix C.2.1 provides specifics on the ACIS data, and Appendix C.3 describes our cleaning process.

A key variable in these data is the outcome of incarceration. A central choice judges make in sentencing is whether to administer a community punishment (fines or probation), an intermediate punishment (probation with additional restrictions), or an active sentence, which entails being incarcerated in prison or jail.⁵ Judicial discretion to impose active sentences is restricted by North Carolina’s structured sentencing system; Appendix B provides institutional background on the North Carolina court and structured sentencing systems.⁶

⁴ See <https://www.census.gov/data/tables/time-series/demo/families/children.html>.

⁵ We use the terms *active sentence* and *incarceration* interchangeably throughout the rest of the paper.

⁶ Criminal defendants who do not receive active sentences may still experience brief spells of incarceration as a consequence of pretrial detention or intermediate sentences. Our analysis compares individuals who face incarceration spells from active sentences against individuals who do not receive as severe of a punishment but may still experience temporary jail spells. This generally biases us against finding an effect of incarceration on other outcomes.

Education Records (NCERDC). The second data source is drawn from longitudinal records provided by the North Carolina Education Research Data Center (NCERDC) that cover all K–12 public school students in the state from 2006 to 2017. Throughout the paper, we define years from July–June (rather than January–December) to align with the school calendar. For example, the year 2010 refers to both defendant and student outcomes measured from July 1, 2009 to June 30, 2010. We observe students’ demographic characteristics and their school and grade in each year. Our main academic outcomes are students’ scores on state standardized math and English exams. Our measure of math achievement includes scores on both grade 3–8 math tests and an end-of-course high school algebra exam. English scores include performance on grade 3–8 reading tests and an end-of-course high school English I exam. We standardize all scores to have a mean of zero and a standard deviation of one in the population of test takers for each exam cohort.

We observe a wide range of behavioral outcomes such as days of absence, suspensions, and fighting incidents. The NCERDC data also include students’ geocoded addresses each year, which facilitates our merge with the court data.⁷ Appendix C.1 defines our key variables and Appendix C.2.2 describes the NCERDC data in detail.

2.2 Address-Based Merge Between Education and Court Records

Our analysis is facilitated by a unique link between the statewide court (ACIS) and education (NCERDC) datasets. We sent the NCERDC a list of the addresses that are available in our court records, which include information on street address, city, state, and zip code. The NCERDC linked these variables to the geocoded student address identifier using confidential information and provided us with a crosswalk between addresses in the two datasets. Using this crosswalk, we link students and defendants who live at the same address in the same year.⁸ Appendix C.4 provides details on the merge of the ACIS and NCERDC datasets.

Our comparison of linked addresses and individuals suggests that the merge is of high quality. We geocoded addresses in the court data to obtain the Census Block Group of each defendant (this variable also appears in the education data). We find that 97 percent of linked addresses are in the

⁷ The education data cover both traditional public schools and charter schools, but NCERDC does not collect addresses from most charter schools. Since our merge of the education and court data relies on addresses, most students in our merged samples are from traditional public schools.

⁸ After we have linked a student and a criminal defendant, we connect the student to all of the defendant’s cases that we observe in the court data regardless of whether their addresses match in other cases.

same Census Block Group in cases where this variable is defined in both datasets (see Appendix Table C1). In addition, we find that 84 percent of linked defendants and students have the same race/ethnicity, even before we impose further sample restrictions (see Appendix Table C2).

Table 1 provides summary statistics on the linked education and court data. Column (A) includes all students who attended a North Carolina public school in 2010–2014. Column (B) restricts to the students who were not linked to a defendant in the court data while column (C) restricts to the students who were linked. Column (D) includes only those students who were linked to a defendant who received an active sentence in these years. Panel A displays mean demographic characteristics for students, Panel B shows means of our main outcome variables, and Panel C shows characteristics of the criminal defendants for matched subsets.⁹ Relative to other public school students, children who are linked to a criminal defendant are more likely to be Black and economically disadvantaged; they also have lower test scores and more misbehavior incidents. These patterns are even more pronounced in the subsample of students who were exposed to an active sentence in their household (column (D)).

Our address-based merge has two important caveats. First, we do not observe the relationship between the student and the defendant. As a consequence, our results reflect the incarceration of any household member. Table 1 shows that the average linked defendant in our data is 33 years old at the time of case disposition, while the average linked student is roughly 12 years old. Thus, many linked defendants are likely to be the child’s parents, but our sample also includes siblings, grandparents, aunts and uncles, and other relations. In some analyses we infer the nature of the relationship using age gaps between defendants and children.

A second caveat is that a merge based on address may create false matches in which the student and defendant are not living together. Since student addresses are recorded only once per year, false matches can occur if families move mid-year.¹⁰ Another source of potential mismatch arises from apartment buildings and other multi-unit addresses. To address this, we impose several restrictions to reduce the rate of false matches in the analysis samples. Appendix C.5 provides details on the construction of our samples for the analyses in Sections 3–5.

⁹ If the student was linked to more than one defendant or offense, Panel C reports statistics for the most serious offense (see Appendix C.5).

¹⁰ The ACIS records update addresses if defendants move, but mismatch could also arise from lags in this updating.

2.3 Baseline Evidence of an Achievement-Incarceration Gradient

Figure 1 shows a raw version of the central empirical relationship of our paper: a large negative correlation between community incarceration rates and student achievement. Panel A displays a negative and virtually linear relationship between (log) school-level exposure to incarceration and math scores for all North Carolina public school students from 2010 to 2014. In this panel, school exposure is the average number of students who experience a household incarceration in a typical year as computed in our full matched sample. Panel B shows a similar negative relationship between census tract-level incarceration events and math scores.

Table 2 shows OLS regressions capturing the same relationship. Incarceration rates are a strong predictor of student achievement at both the school and neighborhood level.¹¹ Accounting for demographics as well as county and year fixed effects reduces the magnitude of the gradient by 50–60 percent, but a robust and economically meaningful relationship remains. The gradient persists when excluding all children with an incarceration in their household or all children matched to court data. This indicates that the negative correlation between incarceration and achievement is not solely driven by families directly involved with the criminal justice system.

Even conditional on controls and fixed effects, these correlations need not reflect causal relationships. Families likely sort to different neighborhoods and schools based on income, job opportunities, crime rates, and other factors associated with local incarceration rates. In the next section, we develop a judicial turnover strategy to isolate the causal effect of community-level incarceration on student achievement.

3 A Judicial Turnover Design

The ideal experiment to estimate the impacts of community-level incarceration on student achievement would be to randomly vary the fraction of criminal cases that result in incarceration across communities. While such an experiment is infeasible in practice, our strategy approaches this ideal by isolating plausibly exogenous variation in county-level incarceration rates. We do so by

¹¹To provide context on magnitudes, a one log point increase in incarceration at the school level is associated with a reduction in both math and English scores of roughly 20 percent of a standard deviation (Panel A). At the neighborhood level, this rises to nearly 30 percent of a standard deviation (Panel B). The average school in our data experiences nearly 17 household incarceration events in a typical year. Thus, a single additional incarceration linked to the mean school is associated with test scores that are 1.2 percent of a standard deviation lower for *all* students.

exploiting two features of the court system: (i) the relatively stable tendency of particular judges to be more or less severe in incarcerating defendants and (ii) turnover of judges in different North Carolina counties during our sample period.

The intuition behind the identification approach is that when a more stringent judge—one with a greater tendency to impose active sentences—enters or leaves a county, we expect aggregate incarceration levels to change. Specifically, we expect incarceration to rise if a stringent judge enters and to fall if a stringent judge departs. We examine turnover from both arrivals and departures, which ultimately stem from judicial elections, mid-term leaves, and pre-scheduled rotations across counties.

3.1 Empirical Strategy and Implementation

We exploit year-to-year changes in county-level average stringency that are driven by the introduction or departure of individual judges. This is similar in spirit to [Chetty, Friedman, and Rockoff \(2014\)](#)'s strategy of using teacher arrivals and departures to validate teacher value-added estimates, and we follow their empirical specification. For county c and year t (and judge j), we define the change in average stringency as

$$\Delta S_{ct} = \sum_j (\omega_{jct} - \omega_{jct-1}) \cdot \mu_{jct}^{-\{t-1,t\}}. \quad (1)$$

The term inside the sum has two components: a leave-out measure of judge j 's stringency, $\mu_{jct}^{-\{t-1,t\}}$, and the change in judge j 's share of the caseload between year $t - 1$ and t , $(\omega_{jct} - \omega_{jct-1})$.

Judicial Stringency $\mu_{jct}^{-\{t-1,t\}}$. Our measure of judicial stringency is intended to capture judge j 's average tendency to impose an active sentence among the convictions they oversee (versus an intermediate or community punishment). There are two concerns with defining stringency using a simple average of the binary outcome of imposing an active sentence. First, individual judges may face caseloads of differing severity. To account for this, we residualize the binary measure with respect to structured sentencing cell fixed effects. Under the structured sentencing system in North Carolina, each case is assigned to a particular cell of a sentencing matrix on the basis of offense severity and the defendant's history of prior points (see Appendix Figures [B1](#) and

B2). Within each cell, judges have discretion over the prescribed range of potential sentences. Conditioning on structured sentencing cells therefore isolates the variation in active sentencing that is attributable to judicial discretion.

The second concern is that judges present in a particular county or year with a large number of relatively serious offenses (within sentencing cell) are more likely to impose active sentences. As a result, a high average active sentence rate might reflect the features of a location or time period rather than a characteristic of the judge. To address this possibility, we compute a leave-out average, again following the approach of [Chetty et al. \(2014\)](#). Specifically, for judge j in county c and year t , we define $\mu_{jct}^{-\{t-1,t\}}$ to be the average value of the active sentencing residual for j leaving out any case in county c in years t or $t - 1$. Similar to a jackknife leave-out mean, this approach gives a measure of the judge’s tendency that is not directly influenced by the specific time and place. For our main estimate of $\mu_{jct}^{-\{t-1,t\}}$, we limit our sample to judge-year pairs with at least 50 observations.

Change in Caseload Weights ($\omega_{jct} - \omega_{jct-1}$). When computing ΔS_{ct} we fix the measure of stringency for each judge, $\mu_{jct}^{-\{t-1,t\}}$. As a result, changes in county-level stringency are driven only by changes in the number of cases that each judge hears in that county, as captured by the caseload weights, ω_{jct} . For our baseline analysis, we measure these weights using judges’ observed caseloads. In other words, we define ω_{jct} to be the fraction of all criminal cases in county c and year t that were heard by judge j . When a judge enters a county, the weight goes from 0 to a positive number. When a judge leaves a county, the weight goes from a positive number to 0.

Because our baseline uses observed caseloads, ΔS_{ct} captures changes in average stringency that arise from the arrival and departure of judges in a county as well as changes in the number of cases heard by each judge within a county. This latter component includes variation in caseloads from vacations, unusually long trials, and other scheduling constraints, and could potentially reflect endogenous assignment of caseloads within-counties. To isolate variation due only to turnover, we include robustness checks that restrict ω_{jct} to vary solely as the result of judges’ entry and exit (see [Section 3.7](#)).

3.2 What Generates Variation in County Level Stringency ΔS_{ct} ?

In general, elections are the most prominent source of turnover in our sample. Appendix Table A1 provides a variance decomposition of ΔS_{ct} . 60 percent of the variation in our baseline measure is due to turnover, and roughly a third of this stems from elections. Mid-term leaves also make up a meaningful fraction of the variation in ΔS_{ct} , while cross-county rotations play a smaller role. Overall, there is substantial variation in the average tendency of judges to impose an active sentence across counties; the standard deviation of county stringency is 7 percentage points (in terms of the probability of assigning an active sentence), and the change from the 10th to 90th percentile is 20 percentage points. Appendix Figures B3 and B4 show the average stringency of district and superior court judges in each county in 2010, along with boundaries for district and superior courts. Appendix B provides details on judicial elections and the sources of judge turnover.

Appendix Table A2 provides examples of judge turnover events that led to the largest changes in county-level judicial stringency (ΔS_{ct}) in our sample. These changes are evenly spread across county sizes and our sample period, and are driven by gubernatorial appointments to replace judges departing mid-term, rotations of judges between district and superior courts, and election results that bring in a new judge. Notably, a non-trivial fraction of the elections, even in these extreme examples, are decided in relatively close races, a fact that we exploit in some of our robustness tests.

3.3 County-Level Stringency Impacts Incarceration Rates

We begin by showing that shifts in county-level stringency lead to large changes in the realized number of active sentences in a county. Panel A of Table 3 shows results from the following regression:

$$\Delta Y_{ct} = \beta \cdot \Delta S_{ct} + \gamma_t + \epsilon_{ct}. \quad (2)$$

Here, $\Delta Y_{ct} \equiv Y_{ct} - Y_{c,t-1}$ and Y_{ct} measures the log number of active sentences in a given county and year.¹² γ_t are year fixed effects and standard errors are clustered at the county level. Across all

¹² The regressions based on Equation 2 are at the county \times year level, but we weight observations by the number of individual-level observations used to compute ΔY_{ct} . When ΔY_{ct} measures log totals (active sentences, criminal cases, or employment), we hold fixed these weights based on the number of individual-level observations in 2010 (the first year in our sample).

specifications, we scale ΔS_{ct} so that one unit corresponds to one standard deviation in the baseline distribution of county stringency (0.07).¹³

The first row of Panel A shows that a one standard deviation increase in county stringency is associated with a 0.189 log point increase in the number of active sentences. Put differently, a one standard deviation increase in county-level stringency generates a 21 percent increase in the annual number of active sentences (approximately 400 incarcerations in the average county).

A dynamic version of this specification, presented in Panel A of Figure 2 and Appendix Table A3, shows that incarcerations jump precisely when changes in stringency occur. This figure plots coefficients from a modified version of Equation 2 that considers longer horizon differences in incarceration. To evaluate pre-trends, we consider changes between $t - 1$ and deeper lags (e.g. $Y_{c,t-3} - Y_{c,t-1}$). To evaluate post-event changes, we consider differences between $t - 1$ and $t + 1$ or $t + 2$. In all cases ΔS_{ct} is fixed as the difference between t and $t - 1$.¹⁴ We do not see evidence of pre-trends: changes in county-stringency are unrelated to past changes in incarceration. Conversely, the number of active sentences rises sharply after S_{ct} increases. This is consistent with harsher judges increasing incarceration rates.

Importantly, this effect is not driven by a contemporaneous increase in arrests or criminal cases. Panel B of Figure 2 repeats our dynamic specification with changes in log criminal cases as the dependant variable. We do not see evidence of pre-trends or of any meaningful change in cases after a change in stringency takes place. We consider balance and potential threats to our strategy in more detail below, but the stability of criminal caseloads rules out a key potential concern: the rise in incarceration shown in Panel A is not an artifact of large increases in crime.

The remaining rows of the first panel of Table 3 show that these changes meaningfully impact the exposure of local students to incarceration. A one-standard deviation increase in county stringency increases the probability a student is directly exposed within the household by 0.3 percentage points (about 13% of the mean). Strikingly, the impact on indirect exposures is sub-

¹³ Note that we scale our coefficients by the standard deviation of S_{ct} (0.07) rather than by the standard deviation of ΔS_{ct} (0.018). This scaling is policy relevant in the sense that it reflects baseline differences in the average tendency to incarcerate criminal defendants across counties.

¹⁴ Specifically, we fix the caseload weights to capture judge turnover between t and $t - 1$. To avoid a mechanical relationship, the underlying individual judge stringency measures that are used to build ΔS_{ct} for Figure 2 are computed leaving out all observations between years $t + k$ and $t - 1$ for each regression, i.e. $\mu_{jct}^{-\{t-1, \dots, t+k\}}$. Results are effectively unchanged if we use the baseline measure of ΔS_{ct} and do not make this adjustment.

stantially higher. For example, the average student experiences a 0.56 increase in the number of classmates who are directly exposed. In other words, increases in judicial stringency lead to substantial growth in both incarceration rates and community-level exposure to incarceration.

3.4 Increases in County-Level Stringency Reduce Local Academic Achievement

We next show that changes in county-level stringency negatively affect the academic performance of local children. Given the robust relationship between ΔS_{ct} and incarceration rates, and the absence of any meaningful relationship with criminal case totals, defendant characteristics, and student demographics (which we discuss in detail below), we interpret this as causal evidence that community-level incarceration adversely affects student achievement.

Panel B of Table 3 presents our main results. These regressions follow the specification in Equation 2 but define ΔY_{ct} as the change in county-average test scores in either math or English. We find negative impacts for both, with larger effects on math scores. Column (B) shows that a one standard deviation increase in county stringency results in a statistically significant decline of 0.024 standard deviations for math scores and a 0.014 standard deviations for English scores. We also find similar results when focusing on within-student changes in test scores, suggesting that these estimates are not driven by compositional shifts in the student population.¹⁵

Dynamic versions of these specifications, shown in Panels C and D of Figure 2 (and in Panel B of Appendix Table A3), indicate that our estimates are not driven by pre-trends in test scores. There is little relationship between ΔS_{ct} and test score growth in years before t , but stringency predicts decreases in test score at time t and afterward (with the magnitude of the coefficients rising slightly over time). In addition to providing evidence against confounding long-run trends, these results suggest that community-level exposure to incarceration has persistent and accumulative negative effects on student achievement.

Assuming the impact of county-level stringency operates *only* through the channel of incarceration, these estimates imply that a large fraction of the neighborhood gradient in Figure 1 reflects a causal relationship. Our turnover estimates in Table 3 suggest that a one log point increase in county-level active sentences leads to a 0.127 standard deviation reduction in math scores.¹⁶ This

¹⁵ Results are available on request and in earlier publicly posted drafts of our paper.

¹⁶ This is derived from a simple Wald-style ratio of the impacts of county stringency on math scores (-0.024) and

represents just over 40 percent of the unconditional gradient shown in Panel B of Table 2 and is nearly identical to the estimated gradient after controlling for observables and fixed effects (see columns (B) and (C)). A similar calculation suggests that the causal effect is responsible for nearly 25 percent of the raw neighborhood-level gradient in English scores. The aggregate relationship between incarceration and achievement does not appear to be an artifact of selection.

3.5 Achievement Falls Even for Students Who Are Not Directly Impacted

The adverse impact of incarceration is not limited to children whose families are directly involved with the criminal justice system. In Columns (C) and (D) of Table 3 we repeat the specifications in column (B) but split the sample into students who are linked versus not linked to a defendant in the court records. Even in the linked sample, a relatively small fraction are exposed to an incarceration in the household—many are linked to defendants who are acquitted, have charges dropped, or receive lesser sentences. As such, the estimated coefficient for this population should not be interpreted as the impact of direct exposure (which we explore in detail below). However, the set of never-linked students provides a relatively clean sample of students who *do not* experience direct exposure to household incarceration.

We find that our estimates for this never-linked sample (column (D)) are sizeable and statistically significant. The coefficients for both math and English scores are similar in magnitude to our baseline.¹⁷ The fact that we observe a large adverse impact on children who do not experience the incarceration of a household member suggests that the aggregate impacts of incarceration must operate, at least in part, through community-level factors that are shared by many children through indirect ties. In other words, that there are substantial community level consequences of incarceration. We explore this possibility in more detail in Sections 4 and 5.

log number of active sentences (0.189). This calculation also assumes linearity in the impacts of county stringency. Appendix Table A4 shows evidence broadly consistent with this assumption; increases in county stringency lead to increases in active sentences and reductions in test scores in communities with both high and low exposure to criminal cases, although these effects are slightly larger in the schools and neighborhoods with the highest exposure.

¹⁷This should not be interpreted as evidence that the impacts are the same for directly and indirectly exposed students, or larger for indirectly exposed students. The coefficients across specifications are statistically indistinguishable, and the fraction of students that are directly exposed to incarceration, even in the linked sample, is small enough that even very large impacts on directly exposed children would not meaningfully impact the average.

3.6 Changes in Stringency Are Not Related to Changes in County-Level Observables

We next address the central challenge to our judge turnover approach: the possibility that changes in judicial stringency correlate with time-varying county-level factors beyond incarceration (which in turn influence educational outcomes). This might happen, for example, because harsh judges are elected as a county trends toward more criminal activity or as county demographics change. We provide evidence against this concern in Table 4. This table shows balance tests based on Equation 2, replacing the outcome with various county-level observables.

Panel A shows that our measure of stringency is unrelated to changes in local criminal activity (expanding on our dynamic results for criminal caseloads presented above). For these specifications, ΔY_{ct} measures changes in different types of criminal cases in a county between years $t - 1$ and t . Column (B) shows that our benchmark measure of ΔS_{ct} is not related to changes in the total number of cases or to the total numbers of cases by offense type (felony, misdemeanor, traffic, or clerk to decide). Judicial stringency does not appear to shift as a consequence of underlying trends in crime or criminal cases. This addresses an important potential confound in the descriptive results in Section 2 as well as in many interdisciplinary analyses of the community incarceration. We present dynamic versions of these specifications in Panel C of Appendix Table A3.

Another plausible driver of judicial stringency is shifts in labor market conditions, which might change attitudes among voters, affect judicial decision-making, or otherwise impact the behavior of criminal defendants. In Panel B, we find little evidence that adverse labor market conditions lead to increases in ΔS_{ct} . If anything, increases in county-level stringency are slightly (but insignificantly) associated with decreases in unemployment. The joint F -statistic across labor market outcomes is 1.7, suggesting that labor markets are unlikely to be a crucial omitted factor confounding our results. We present the results of dynamic versions of these specifications in Panel D of Appendix Table A3.

We also find that ΔS_{ct} is balanced with respect to demographic changes. We do not find associations between changes in stringency and changes in the average characteristics of criminal defendants or local students. Panel C shows that ΔS_{ct} is unrelated to changes in the average race, age, gender, or criminal history of defendants at the county level. Panel D shows that ΔS_{ct} is not related to changes in students' age, gender, or socioeconomic status; here we find marginally

significant associations with student race, but the point estimates are small. The F -statistics from tests of joint significance across all defendant or student characteristics range from 1.0–1.5, suggesting that we cannot reject random assignment. We find precisely estimated zeroes for *predicted* math/English scores based on a large vector of student characteristics (age, gender, race, economic disadvantage, English proficiency, and disability codes). As a whole, the results in Panels A–D of Table 4 indicate that judicial turnover is, at least on average, unrelated to changes in observable factors that impact students’ educational outcomes.

3.7 Robustness: Alternative Measures of Stringency

Our evidence so far suggests that incarceration causally reduces educational achievement in local communities, including children who are not directly affected by a household incarceration. Table 5 shows that our results are robust to different sources of variation in county-level stringency. For reference, column (A) in this table replicates our benchmark results on active sentences and test scores (from column (B) of Table 3). The other columns define alternative versions of ΔS_{ct} to isolate and focus on different sources of variation in stringency. We now discuss these in turn.

Similar Results when Focusing Strictly on Entry and Exit. To capture changes in stringency that occur only because of judges entering or exiting a county, we consider a version of ΔS_{ct} that fixes the weights, ω_{jct} , for all years that judge j is active in county c . We refer to this as our *fixed caseloads* measure. For this measure, ω_{jct} is equal to zero if judge j is not active in county c in year t , and is otherwise equal to judge j ’s average fraction of cases heard in county c across all active years in our sample.¹⁸ With the fixed caseload measure, variation in stringency comes only from turnover in judges, which results from either (i) elections, (ii) mid-term departures (due to retirement or appointment to higher courts), or (iii) rotations across counties. Column (B) of Table 5 shows that we continue to find negative effects on math and English scores in this approach, with slightly larger estimates than in our baseline specification.

Column (C) in Table 4 shows that our fixed caseload measure is also balanced with respect to criminal activity (Panel A), employment conditions (Panel B), and defendant and student demographics (Panels C–D). Joint F -statistics range from 1.0 to 1.3 across these four sets of outcomes.

¹⁸ We consider judge j to be active in county c if they preside over at least 20 cases.

This suggests that the entry and exit of more stringent judges is unrelated to changes in observable factors that correlate with student achievement.

Results are not the Mechanical Consequence of Including Future Decisions in ΔS_{ct} . Our baseline measure of judicial stringency leaves out data from years t and $t - 1$. As a consequence, stringency for new judges is measured based on future cases.¹⁹ A possible concern is that using future data to measure contemporaneous stringency might lead to biased estimates (for example, because future incarceration rates reflect emergent trends in the quantity or unobserved severity of criminal activity). To ensure that this is not driving our result, column (C) of Table 5 focuses on variation in stringency that comes only from *departing* judges. Specifically, we fix $(\omega_{jct} - \omega_{jct-1})$ to be 0 unless judge j exits the county in period t (in which case ω_{jct} is set to 0 and ω_{jct-1} is set to j 's average caseload in c while active, as in our fixed caseload measure). As a result, the county level change in stringency is determined solely based on judge decisions made prior to time t . We find similar effects in this specification, with slightly larger negative effects on math and English scores than in our benchmark specification.

Results are not Driven by Strategic Voting. Because elections are a major source of variation in stringency, a natural question is whether our findings are similar when considering *only* election driven turnover. In column (D) of Table 5 we repeat our fixed caseloads measure, but focus strictly on entry and exit that comes as a result of elections (i.e. we fix $(\omega_{jct} - \omega_{jct-1})$ to be 0 unless judge j leaves or enters c because of an election). The results are in line with our benchmark approach.

Of course, elections may, in principle, be a confound in our analysis. For example, voters may strategically elect certain types of judges in response to past (or coming) trends in crime or student achievement. We present four pieces of evidence against this concern. First, column (E) of Table 5 shows that estimates are similar when focusing on variation that comes only as the result of close elections—those in which the winning judge's share of the vote was below 55 percent. In these elections there was no clear mandate for one judge over another and the outcome could feasibly have gone in either direction. Given this, the resulting variation is plausibly close to random, and less likely to be related to future changes in crime or student achievement.

¹⁹Our approach follows [Chetty et al. \(2014\)](#) in using both past and future years to increase power.

In a typical election, voters may signal disapproval with a sitting judge (due to crime or sentencing policies) by voting them out, which may generate a correlation between stringency and our outcomes that does not run through incarceration. To address this potential concern, our second piece of evidence, in column (F) of Table 5, focuses on variation only due to retirement (which tends to be the result of age). All judges running for election are new to voters in these cases, so the change in stringency is less likely to be related to the characteristics of the departing judge. We again find similar results.

Third, all judge elections during our period were non-partisan, so voters may have had little information about which judges were likely to be stringent. To bring home this point, Appendix Figure B5 shows that there is essentially no relationship between ΔS_{ct} and changes in county-level Republican vote shares in the 2010 and 2012 elections.

Fourth, and finally, column (D) of Appendix Table A1 shows that results are similar, and in fact larger, when focusing only on variation that is not election driven. Specifically, we focus on entry and exit that is due only to mid-term departures, which stem from retirement, appointment to higher-courts, and other factors. In some cases these judges are replaced by political appointees and in others the position goes temporarily unfilled. This confirms that our findings are not strictly due to elections. Collectively, these tests help to rule out the possibility that voter sentiment (potentially due to underlying crime rates) is driving changes in stringency.

4 Mechanisms: The Classroom Disruption Channel

The results in Section 3 indicate that aggregate incarceration rates adversely impact the academic achievement of entire communities. A crucial next question is how this effect occurs. The magnitude of our estimates, and the fact that they persist in a sample of students who are not directly linked to court records, indicates that a large fraction of the relationship between incarceration and test scores must be due to spillovers onto a broad set of children. While there are many plausible channels through which such a widespread impact may operate, we focus on one particular mechanism: spillovers from directly exposed students onto their classmates (inspired by the literature on classroom disruptions). If exposure to the incarceration of a household member impacts a student's misbehavior at school, we might expect this to disrupt a large number of other children

within the classroom. In this section we test whether direct exposure to household incarceration leads to disruptive behavior within the classroom.

4.1 Empirical Specifications

OLS Estimation

The object of interest in this section is the causal impact of direct exposure to an active sentence within the household. As a baseline, we focus on a sample of students that are linked to criminal defendants at the address level and consider a simple OLS specification. While this approach is likely to provide biased estimates, it is a useful benchmark to compare against better identified approaches (and allows us to introduce notation). For student i in school year t we have:

$$y_{it} = \beta \cdot \text{Active Sentence}_i + \theta_{oc\tau(i)} + \mathbf{X}_{it}'\zeta + v_{it}. \quad (3)$$

y_{it} represents student i 's outcome measured in or after the year of the defendant's criminal disposition. Active Sentence_i is our measure of incarceration, which is an indicator equal to one if student i 's household member received an active sentence.²⁰ $\theta_{oc\tau(i)}$ is a fixed effect for offense class at arrest (o) \times court (c) \times disposition year (τ), each of which are characteristics of the defendant's case. \mathbf{X}_{it} is a vector of student and defendant covariates.²¹ For these regressions, we consider all students matched to court records except for (i) those linked to low-level traffic offenses and (ii) likely false matches. We also drop cases that are missing information on a judge (or matched to a judge who appears rarely in the data) and consider only the most serious charge associated with each student. Appendix C.5 provides details on the construction of the sample and Appendix Table A5 shows summary statistics.

There are two main endogeneity concerns with this cross-sectional OLS regression. First, it compares convicted defendants to those who are charged but not convicted. As a result, it risks conflating the traumatic impacts of the offense itself (or the circumstances precipitating the of-

²⁰ Note that many of our variables, including Active Sentence_i , are characteristics or outcomes of the *defendant* linked to student i rather than to student i themselves. We define our direct exposure sample so that each student maps to only one case and one defendant, and so we use i as the subscript for these variables to reduce notation.

²¹ We define courts as a county \times court type (district or superior) pair. Throughout our analysis, we define offense controls at the *class* level (e.g., felony F, misdemeanor 2). \mathbf{X}_{it} includes defendant age, gender, and race; student age, gender, race, and economic disadvantage; and indicators for missing values of each variable.

fense) with the effects of incarceration. Second, incarceration status is likely correlated with unobservable characteristics of defendants and students that impact achievement. To address these issues, we adopt two complementary strategies from the literature: an event study approach and an individual-level judicial stringency approach.

Event Studies

The event study approach addresses the above identification concerns by comparing outcomes for individual students before and after they are exposed to an incarceration. This allows us to include individual-level fixed effects that account for time-invariant student- or defendant-level unobservables. In addition, to avoid conflating the effects of the offense with the impacts of incarceration, we include a control group consisting of similar students with household members who are *convicted* of a similar offense but not incarcerated (because they received probation, fines, or other non-active punishments). For student i in school year t , we estimate:

$$y_{it} = \delta \cdot \mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i + \eta_i + \lambda_{og\tau(i)t} + \varepsilon_{it}. \quad (4)$$

η_i is an individual fixed effect and $\lambda_{og\tau(i)t}$ is a granular set of offense class (o) \times academic cohort (g) \times disposition year (τ) \times school year (t) fixed effects.²² The regressions include all years t in which we observe student outcomes, and our variable of interest is the interaction between Active Sentence_i and an indicator for years of, or after, the case disposition, $\mathbb{1}\{t \geq \tau(i)\}$. Intuitively, the fixed effects restrict identification to students who are in the same graduation cohort and whose household members were convicted in the same year of similar offenses. The coefficient of interest, δ , is a weighted average of the effects of a household incarceration within these covariate cells. Equation 4 is a “stacked regression” that aligns treatment and control observations in event time, which avoids concerns about staggered designs (as in [Cengiz, Dube, Lindner, and Zipperer, 2019](#) and discussed in [Baker, Larcker, and Wang, 2021](#)).

This event study requires a slightly different sample relative to our OLS regressions; namely we (i) include observations before the disposition year, (ii) restrict to students matched to *convicted* defendants, and (iii) select the most serious-linked conviction (rather than charge) for each

²² We define a student’s academic cohort, g , as their year of *expected* high school graduation assuming on-time progression. This is based on the student’s grade level in the first year they appear in the NCERDC data.

student. Appendix C.5 provides details on this sample’s construction, and Appendix Table A5 shows summary statistics.

Judicial Stringency Design

Our event study approach requires a parallel trends assumption. This might fail, for example, if incarceration (conditional on conviction) is related to time-varying circumstances within the household. As an alternative, our judicial stringency approach exploits the quasi-random assignment of defendants to judges of different stringency levels *within* a court and offense class. We thus focus on students linked to defendants who differ in the likelihood of receiving an active sentence only because of the particular judge they faced.

We estimate a reduced form specification that mimics Equation 3 but replaces Active Sentence_{*i*} with $\mu_{j(i)}$, the stringency of the judge in question.²³

$$y_{it} = \gamma \cdot \mu_{j(i)} + \theta_{oct(i)} + \mathbf{X}_{it}'\zeta + u_{it}. \quad (5)$$

The coefficient of interest, γ , measures the impact of a household member being assigned to a more stringent judge. Because the likelihood of an active sentence increases for those assigned to more stringent judges, this provides a reduced form estimate of the causal impact of incarceration. We also estimate an IV specification in which Equation 5 serves as a first stage (with Active Sentence_{*i*} on the left-hand side) and Equation 3 serves as a second stage.

The key identification assumption is that judge stringency, $\mu_{j(i)}$, is unrelated to students’ potential outcomes, conditional on controls. The assignment of cases to judges in North Carolina, while not explicitly randomized, is quasi-random; this reflects the pre-determined allocation of judges to courtrooms at the discretion of the senior judge in consultation with county clerks and the assignment of defendants across courtrooms. The frequent movement of judges across counties

²³ This stringency is defined as a residualized leave-out mean at the judge level, which we compute using only criminal defendants in our data who are *not* linked to any student. Specifically, we regress an active sentence indicator on structured sentencing cell fixed effects in this leave-out sample, and then compute the mean of the residuals at the judge level. The only difference between our measures of judge stringency in Section 3, $\mu_{jct}^{-\{t-1,t\}}$, and Section 4, μ_j , is the set of observations that we exclude when constructing each measure. In Section 3, we leave out specific counties and years. In Section 4, we exclude all defendants who are linked to any student in our data. Using an entirely separate sample to compute stringency avoids any mechanical relationship between μ_j and the disposition for the case linked to *i*, regardless of the set of fixed effects included. We obtain similar results if we leave-out only the defendant in question. See Appendix C.1 for details on our judge stringency measures.

and courtrooms, in both district and superior courts, ensures that judges oversee a variety of offense types. Indeed, balance tests support the assumption that stringency is effectively randomly assigned. $\mu_{j(i)}$ is not correlated with characteristics of the defendant, offense, or cohabitating children but is strongly related to the probability of receiving an active sentence (see Panel A of Table 6, the first stage F-statistic is 103.0). Appendix B discusses the North Carolina judicial process in more detail and presents these balance tests.

4.2 Results: Consequences of Direct Exposure to Incarceration

Across our empirical strategies, we find consistent evidence that direct exposure to incarceration within the household has adverse impacts on student behavior as well as academic performance.

Misbehavior and Attendance

Panel B of Table 6 shows evidence that direct exposure to incarceration leads to increased misbehavior in school. Our event studies—which focus on the impact of incarceration among those exposed to a convicted household member—are in line with basic OLS approach. These results indicate that an incarceration in the household leads to 0.25 additional days suspended, an 8 percent increase in the probability of being suspended at all, and a 4 percent increase in the probability of being disciplined for fighting. We also see an increase in absences of roughly 0.5 days for treated students.

We also consider a dynamic version of our event studies to provide supporting evidence that these effects are not driven by differential trends between treated and control students. We modify Equation 4 by replacing the $\mathbb{1}\{t \geq \tau(i)\}$ term with dummies for years relative to the defendant’s disposition (omitting the year prior to disposition). In Figure 3 we plot the coefficients on these interaction terms from three years before to three years after conviction.²⁴

²⁴ Specifically, we plot δ_l coefficients from the specification

$$y_{it} = \sum_{l=-8}^7 \delta_l \cdot \mathbb{1}\{t - \tau(i) = l\} \times \text{Active Sentence}_i + \eta_i + \lambda_{og\tau(i)t} + \varepsilon_{it},$$

where $l = -1$ is the omitted baseline year. Note that years relative to disposition, $l = t - \tau(i)$, are subsumed by the $\lambda_{og\tau(i)t}$ fixed effects. The range of l from -8 to 7 represents the widest possible window in our data; the education data range from 2006–2017, and disposition years range from 2010–2014. We plot only coefficients from $l = -3$ to 3 because estimates are noisy beyond this range.

Panel A of Figure 3 shows sharp evidence that direct exposure leads to increases in suspensions, a key measure of serious behavioral issues at school. We see little evidence of pre-trends and a sizeable increase in suspensions that coincides with the disposition and persists for several years. We similarly see no significant pre-trends in fighting incidents in Panel B of Figure 3, although there is less pronounced evidence of a treatment effect in this graphical representation.

Our judge stringency approach, which does not depend on a parallel trends assumption, reinforces the evidence in our event studies. The reduced form is scaled in different units, but indicates that one standard deviation increase in judge stringency raises the probability of suspension by 0.3 percentage points (2.4 percent of the mean) and leads to marginally significant increases in absences and fighting. Our IV estimates are qualitatively in line with the event study, though larger in magnitude.

Academic Performance

In Panel C of Table 6 we show evidence that an incarceration in the household also has adverse consequences for math and English test scores. Our event study approach indicates that direct exposure reduces math scores by 1.5 percent of a standard deviation and English scores by 1.2 percent. Both are statistically significant. Dynamic versions of these specifications are shown in Panels C and D of Figure 3. We again see minimal pre-trends, and evidence that test-score declines align with the disposition and persist for years afterwards, particularly in math.²⁵

Robustness: Event Studies

A potential concern in our event study approach is that incarceration is not random, conditional on conviction, even within a specific offense class. Panel A of Appendix Table A6 provides evidence that this concern is unlikely to alter our main findings. For these specifications, we layer in granular controls for the defendant's offense (o) and prior points to examine the importance of other time-varying factors that may be related to the severity of the criminal activity or arrest. Column (A) shows results with *no* offense controls, column (B) controls for offense class at the

²⁵We examine heterogeneity in student outcomes in Appendix Figure A1. The relationship between household incarceration and academic performance is particularly large for Black students, who see a decline in test scores of between 2–4 percent of a standard deviation following the incarceration of a household member according to our event study specifications. We also estimate large effects for defendants who are likely to be mothers (female defendants who are 20–40 years older than the student).

time of arrest, and column (C) controls for offense class at the time of conviction (our benchmark specification). Column (D) controls for more granular four-digit offense codes at conviction, and column (E) controls for four-digit offense codes plus a time trend in the defendant's prior points. The magnitudes of the event study estimates decline slightly across columns, but they remain significant and economically meaningful for most outcomes even when including detailed offense characteristics. The fact that our coefficients are relatively stable as we vary the granularity of our controls for offense class (and criminal backgrounds) indicates that unobserved differences in the gravity of offense is unlikely to be the source of our findings.

Robustness: Judge Stringency

Panel A of Appendix Table A7 shows that our judge stringency estimates are robust to different methods of computing stringency (columns (B)–(C)), to including more granular controls for offenses (column (D)), and to examining effects only in the sample of defendants charged with felonies (column (E)). Appendix Figure A2 displays non-parametric versions of our judge stringency estimates. The figure shows that stringency predicts receipt of an active sentence (Panel A), is uncorrelated with prior offenses (Panel B), and predicts math and reading scores (Panels C and D).

4.3 Relationship with the Literature on Direct Impacts of Incarceration

This evidence indicates that there are meaningful adverse consequence of household exposure to incarceration. This contributes to a growing literature on the direct impacts of family incarceration on children's outcomes (e.g., Cho, 2009a; Bhuller et al., 2018; Dobbie et al., 2018; Arteaga, 2021). While papers in this literature vary in whether they find positive or negative impacts across contexts, outcomes, and empirical strategies, we draw on Murray and Farrington (2008), who emphasize the multidimensional impacts of incarceration, to provide a potential reconciliation of varied results. The incarceration of a family member may have adverse consequences in some domains while leading to improved outcomes in others (e.g. because of the removal of a criminal role model).

More generally, we view our findings as consistent with the most similar studies in a U.S. con-

text. The adverse effects we estimate likely co-exist with the positive impacts emphasized, for example, in [Norris et al. \(2021\)](#). To underscore this point, Appendix Table A8 analyzes two key outcomes from that study.²⁶ Our estimates are in line with the conclusions of this prior work. We find positive effects on neighborhood quality for directly impacted students, suggesting that families move to better neighborhoods following a household incarceration. We find negative, but insignificant, estimates for criminal charges, suggesting that these students are at least no more likely to commit offenses themselves (although child offenses are infrequent given our data structure). Finally, the impacts on test scores that we find lie within the confidence intervals proposed in much of this work (including [Norris et al., 2021](#)), although our estimates are more precise.

5 Discussion: Impacts on Classmates and the Community

Our judge turnover strategy suggests that increases in incarceration have adverse consequences for average test scores in the community. The negative effects are present even among children that have no direct connection to the criminal justice system in their households, so they must operate in part through spillovers onto the wider community.

Following [Lazear \(2001\)](#)'s seminal model of classroom disruptions, there is now a large body of empirical evidence showing that students earn lower test scores when they are in the same school cohorts and neighborhoods as children who are prone to misbehavior ([Figlio, 2007](#); [Aizer, 2008](#); [Fletcher, 2010](#); [Neidell and Waldfogel, 2010](#); [Lavy, Paserman, and Schlosser, 2012](#); [Carrell et al., 2018](#); [Billings and Hoekstra, 2019](#); [Billings, Deming, and Ross, 2019](#)). Given our evidence that direct exposure to incarceration increases misbehavior (shown in Section 4), we hypothesize that classroom disruptions caused by directly impacted children are a meaningful driver of the community level relationship.

Identifying spillovers is challenging across contexts, and our sample is not designed to precisely measure relationships between students. However, as a final empirical exercise, we provide suggestive evidence that disruptions generated by directly exposed students have detrimental academic consequences for classmates. We conclude by contextualizing the magnitude of these

²⁶ We consider neighborhood socioeconomic status percentile and the probability that the directly impacted student themselves has a criminal charge. Of course, this comparison is limited by the fact that we only observe outcomes until the child is in grade 12, as opposed to outcomes in young adulthood. See Appendix C.1 for details on how we define these variables in our data.

spillovers with respect to the literature and the baseline correlations shown in Section 2.

5.1 Empirical Specifications

To examine classroom spillovers, we use the same empirical strategies as in Section 4 but focus on students $k \in \mathcal{K}(i)$ who were classmates with some child i in our direct exposure samples. For each directly exposed child i , we define the set of classmates, $\mathcal{K}(i)$, as the students who were in the same school and grade as child i in the year of the defendant’s disposition. We refer to these students as the “classmates” of child i because they tend to move through the school system together and thus frequently share classrooms. Appendix C.5 provides details on this indirect exposure sample, and Appendix Table A5 displays summary statistics.

Our object of interest is the impact of the incarceration of a child i ’s household member on the academic achievement of their classmate k . We conduct each of our three empirical strategies—OLS, event study, and judge stringency—in the indirect exposure sample. The regression specifications are similar to those in Section 4, but the observations are defined at the classmate (k) \times directly exposed child (i) \times school year (t) level. For example, our event study regression for indirect impacts is

$$y_{kit} = \delta \cdot \mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i + \eta_{ki} + \lambda_{\text{og}\tau(i)t} + \varepsilon_{kit}. \quad (6)$$

This specification is nearly identical to Equation 4, but observations are at the kit level, and we include a fixed effect, η_{ki} , for each classmate \times child pair. All other covariates are still defined by the defendant linked to child i . Similarly, our OLS and judge stringency specifications are generalizations of Equations 3 and 5 to the kit level.²⁷

Sample sizes are much larger in our indirect analyses because there are more classmates than directly impacted children. Further, most classmates k appear in our regression samples multiple times because they are linked to more than one child i in our direct exposure samples. To address

²⁷ Specifically, our OLS and reduced-form judge stringency specifications for indirect impacts are

$$y_{kit} = \beta \cdot \text{Active Sentence}_i + \theta_{\text{oc}\tau(i)} + \mathbf{X}_{\text{kit}}' \zeta + v_{kit}, \quad (7)$$

$$y_{kit} = \gamma \cdot \mu_{j(i)} + \theta_{\text{oc}\tau(i)} + \mathbf{X}_{\text{kit}}' \zeta + u_{kit}. \quad (8)$$

The only difference from Equations 3 and 5 is that the covariate vector, \mathbf{X}_{kit} , is defined by the characteristics of the classmate, k , rather than by the directly impacted child, i (both include characteristics of the defendant linked to child i).

these repeat observations and the possibility of correlation in the errors, we cluster standard errors at the school level in all regressions.

5.2 Results: Consequences of Indirect Exposure to Incarceration

Across our empirical strategies, we consistently find evidence that indirect exposure to household incarceration reduces children’s academic achievement, particularly for math scores. Table 7 presents our results. This table is similar in structure to Table 6: column (B) reports OLS estimates, column (C) reports event study estimates, and columns (D)–(E) report judge stringency estimates (reduced form and IV). In column (B) of Table 7, the OLS coefficients show a negative correlation between indirect exposure and test scores, although these estimates are subject to the usual concerns about omitted variable bias in cross-sectional comparisons.

Column (C) of Table 7 shows negative and precisely estimated indirect impacts on test scores in the event study specification. The point estimates imply that the incarceration of a child’s household member lowers their classmates’ math and English scores by 0.4 and 0.3 percent of a standard deviation, respectively. The magnitude of these estimates are reasonable given past literature on classroom spillovers, as we discuss further below. Figure 4 shows graphical evidence supporting these event studies, plotting the same dynamic specification shown in Figure 3, but with the pair $\{k, i\}$ as the unit of analysis. For math scores we see flat pre-trends prior to the disposition, and sizable declines afterwards that persist through three years. The plot for English scores is less stark, as there is a marginally significant decline prior to the disposition and a noisier pattern afterward.

Our judge stringency strategy is less well-powered to detect the relatively small magnitudes of these indirect impacts, but the point estimates are consistent with our other approaches. The coefficients in column (D) of Table 7 imply that a one standard deviation increase in the stringency of the judge linked to the directly impacted child reduces that test scores of that child’s classmates by 0.1 percent of a standard deviation in both math and English. The IV estimates in column (E) of Table 7 imply indirect impacts of a household incarceration of -1.7 percent of a standard deviation in math and -4.7 percent of a standard deviation in English. Furthermore, the assumption that stringency is as-good-as randomly assigns appears to hold in this sample. There is no significant

correlation between our measure of stringency and the observable characteristics of students or defendants (a joint F-statistic of 1.2 for student characteristic and 1.3 for defendant characteristics). This is true despite a large correlation between observables and the probability a defendant linked to a classmate is incarcerated (see Appendix Table B1).

Robustness and Heterogeneity

We next consider a set of robustness tests that follow our analysis for directly affected students. In Panel B of Appendix Table A6, we show that our event study results are consistent across alternate controls for the defendants offense and prior points. This consistency provides evidence that unobserved offense severity is not correlated with time-variation in the test scores of classmates. In Panel B of Appendix Table A7, we show that our judicial stringency approach is robust to alternative definitions of stringency. Appendix Figure A3 shows heterogeneity across different groups. Most notably, we observe larger indirect impacts for Black students in both our event study and judge stringency approaches.

5.3 Magnitudes

The magnitudes of our estimates appear reasonable given literature on within-classroom spillovers. Carrell and Hoekstra (2010) estimate that, in a typical classroom, adding one child who has been exposed to domestic violence lowers other students' math and reading scores by 0.025 standard deviations, which is 18 percent of the direct impacts of exposure to domestic violence. Similarly, we find that the ratio of indirect-to-direct impacts of incarceration on test scores ranges from 10 to 27 percent, depending on the subject and empirical approach (event study versus judge stringency). The magnitudes of our indirect event study coefficients (-0.004 and -0.003) are more than five times smaller than Carrell and Hoekstra's main peer coefficient (-0.025). This indicates that the indirect effects of a household incarceration are meaningful, but not as substantial as the per-student consequences of domestic violence.

5.4 Unpacking the Aggregate Gradient

In this section, as a final step, we provide a back of the envelope quantification of the importance of the classroom disruption channel for aggregate gradient shown in Figure 1. Figure 5 shows a graphical representation of the exercise, with supporting calculations provided in Appendix Table A9. The sum of the shaded areas represents the raw gradient between math scores and school-level incarceration exposure, as in Panel A of Figure 1. This raw gradient implies that, at the mean level of school exposure, one additional household incarceration is associated with a decline in math scores of 1.2 percent of a standard deviation for *all* students. We decompose this raw gradient using the direct and indirect math score estimates from our event study approach (column (C) in Tables 6 and 7) as well the effects of adding demographic controls (column (C) in Table 2).²⁸

The thin blue area in Figure 5 shows that the direct channel accounts for just 0.2 percent of the raw gradient between math scores and school-level exposure to incarceration. In our sample, only 2.3 percent of children are directly exposed to a household incarceration in a typical year. While we find significant negative effects on math scores for these children (1.5 percent of a standard deviation), there are simply not enough children to account for much of the aggregate relationship.

By contrast, the indirect channel can explain 9 percent of the raw achievement-incarceration gradient in math. This greater magnitude is the result of the relative frequency with which students are indirectly exposed to household incarcerations through their classmates. In our data, the typical child is indirectly exposed to 4.3 household incarcerations per year (see Appendix Figure A4). Over 75 percent of public school children in North Carolina have at least one classmate with a household incarceration in a typical year, and over 15 percent of children experience ten or more indirect exposures.²⁹ Thus, while the indirect effects on math scores are small in magnitude (0.4 percent of a standard deviation), they aggregate to explain a large fraction of the average impact. Appendix Table A9 similarly shows that our indirect event study estimates for English scores can explain 6 percent of the raw gradient in English.

Our analysis of mechanisms focuses only on indirect exposures through children in the same

²⁸ For this decomposition, we assume constant effects at the mean level of incarceration exposure, and we ignore any dynamic effects. See Appendix Table A9 for details.

²⁹ There are racial disparities in indirect exposure to incarceration; in our data, the typical Black student has 5.4 directly impacted classmates, while the typical white student has 3.7 such classmates.

school and grade, which is likely just one of many causal channels that contribute to the raw gradient. Indeed, after accounting for spillovers onto classmates and a range of controls, our decomposition suggests that 42 percent of the raw gradient remains unexplained (see Figure 5 and column (C) of Table 2). This is perhaps unsurprising: children interact across grade levels and outside of school and may directly come in contact with incarcerated adults outside of their households. We hope future research will shed further light on other mechanisms underlying indirect community-level impacts. Still, our results collectively suggest that the direct impacts of a household incarceration on children—a key focus of prior research—are important for aggregate achievement primarily because they lead to widespread spillovers within the community.

6 Conclusion

The central contribution of our paper is establishing that the negative impacts of incarceration have a broader reach into local communities than has been conventionally assumed. We estimate the causal effects of community-level incarceration rates on the achievement of local children using a judicial turnover design. We demonstrate that these community effects must be, in large part, the result of spillovers onto children who are not directly exposed to incarceration within their households.

We then highlight an important mechanism underlying this result: behavioral disruptions in the classroom. We show that direct exposure to the incarceration of a household member negatively impacts a student’s misbehavior and test scores, which in turn appear to adversely affect the academic achievement of their classmates. The impacts on achievement are persistent, lasting at least three years at both the individual and aggregate levels. Though small in magnitude for any individual child, these spillovers account for a meaningful share of the overall gradient between achievement and community incarceration rates (although a substantial portion of the gradient is likely driven by other channels). The strength of this indirect channel reflects the fact that children are so frequently exposed to incarceration at school and in their neighborhoods.

While the results are consistent with a large body of descriptive work that has documented how incarceration may negatively impact incarcerated individuals and their family members, this paper advances the literature with a research design that allows us to establish the causal conse-

quences of community-level incarceration rates. Our findings point to the value of criminal justice reform that takes into account the broader dimensions of mass incarceration. Incorporating the spillover effects of incarceration into policy design may help to improve the opportunities of children in some of the nation's most underprivileged communities.

References

- Aizer, A. (2008). Peer effects and human capital accumulation: The externalities of ADD. NBER Working Paper No. 14354.
- Arteaga, C. (2021, 10). Parental Incarceration and Children's Educational Attainment. *The Review of Economics and Statistics*, 1–45.
- Baker, A., D. F. Larcker, and C. C. Wang (2021). How much should we trust staggered difference-in-differences estimates? *Working Paper*.
- Bhuller, M., G. B. Dahl, K. V. Loken, and M. Mogstad (2018). Intergenerational effects of incarceration. In *AEA Papers and Proceedings*, Volume 108, pp. 234–40.
- Billings, S. B. (2018). Parental arrest and incarceration: How does it impact the children? *Available at SSRN 3034539*.
- Billings, S. B., D. J. Deming, and S. L. Ross (2019). Partners in crime. *American Economic Journal: Applied Economics* 11(1), 126–50.
- Billings, S. B. and M. Hoekstra (2019). Schools, neighborhoods, and the long-run effect of crime-prone peers. Technical report, National Bureau of Economic Research.
- Carrell, S. E., M. Hoekstra, and E. Kuka (2018). The long-run effects of disruptive peers. *The American Economic Review* 108(11), 3377–3415.
- Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics* 2(1), 211–228.

- Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics* 134(3), 1405–1454.
- Chetty, R., J. N. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2018). The opportunity atlas: Mapping the childhood roots of social mobility. Technical report, National Bureau of Economic Research.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593–2632.
- Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics* 129(4), 1553–1623.
- Cho, R. M. (2009a). The impact of maternal imprisonment on children’s educational achievement results from children in chicago public schools. *Journal of Human Resources* 44(3), 772–797.
- Cho, R. M. (2009b). Impact of maternal imprisonment on children’s probability of grade retention. *Journal of Urban Economics* 65(1), 11–23.
- Clear, T. R. (2008). The effects of high imprisonment rates on communities. *Crime and justice* 37(1), 97–132.
- Clear, T. R., D. R. Rose, E. Waring, and K. Scully (2003). Coercive mobility and crime: A preliminary examination of concentrated incarceration and social disorganization. *Justice Quarterly* 20(1), 33–64.
- Dobbie, W., H. Grönqvist, S. Niknami, M. Palme, and M. Priks (2018). The intergenerational effects of parental incarceration. Technical report, National Bureau of Economic Research.
- Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education finance and policy* 2(4), 376–394.
- Finlay, K., M. Mueller-Smith, and B. Street (2023). Children’s indirect exposure to the US justice system: Evidence from longitudinal links between survey and administrative data. Working paper.

- Fletcher, J. (2010). Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *Journal of policy analysis and management* 29(1), 69–83.
- Foster, H. and J. Hagan (2007). Incarceration and intergenerational social exclusion. *Social Problems* 54(4), 399–433.
- Frandsen, B. R., L. J. Lefgren, and E. C. Leslie (2019). Judging judge fixed effects. NBER Working Paper No. 25528.
- Geller, A., C. E. Cooper, I. Garfinkel, O. Schwartz-Soicher, and R. B. Mincy (2012). Beyond absenteeism: Father incarceration and child development. *Demography* 49(1), 49–76.
- Hagan, J. and R. Dinovitzer (1999). Collateral consequences of imprisonment for children, communities, and prisoners. *Crime and justice* 26, 121–162.
- Haskins, A. R. (2014). Unintended consequences: Effects of paternal incarceration on child school readiness and later special education placement. *Sociological Science* 1, 141.
- Larson, R., S. Shannon, A. Sojourner, and C. Uggen (2021). Felon history and change in us employment rates. *Social Science Research*, 102649.
- Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal* 122(559), 208–237.
- Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics* 116(3), 777–803.
- Lynch, J. P. and W. J. Sabol (2004). Effects of incarceration on informal social control. *Imprisoning America: The social effects of mass incarceration*, 135–164.
- Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Working Paper* 18.
- Murray, J. (2013). The effects of imprisonment on families and children of prisoners. In *The effects of imprisonment*, pp. 462–482. Willan.

- Murray, J. and D. Farrington (2005). Parental imprisonment: effects on boys' antisocial behaviour and delinquency through the life-course. *Journal of Child Psychology and psychiatry* 46(12), 1269–1278.
- Murray, J. and D. P. Farrington (2008). The effects of parental imprisonment on children. *Crime and justice* 37(1), 133–206.
- Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics* 92(3), 562–576.
- Norris, S., M. Pecenco, and J. Weaver (2021, September). The effects of parental and sibling incarceration: Evidence from ohio. *American Economic Review* 111(9), 2926–63.
- Pattillo, M., B. Western, and D. Weiman (2004). *Imprisoning America: The social effects of mass incarceration*. Russell Sage Foundation.
- Rose, D. R. and T. R. Clear (1998). Incarceration, social capital, and crime: Implications for social disorganization theory. *Criminology* 36(3), 441–480.
- Spainhour, W. and S. Katzenelson (2009). Structured sentencing training and reference manual. *Raleigh, NC: North Carolina Sentencing and Policy Advisor Commission*.
- Thomas, A. and I. Sawhill (2005). For love and money? the impact of family structure on family income. *The Future of Children*, 57–74.
- Wildeman, C. (2010). Paternal incarceration and children's physically aggressive behaviors: Evidence from the fragile families and child wellbeing study. *Social Forces* 89(1), 285–309.
- Wildeman, C. and K. Turney (2014). Positive, negative, or null? the effects of maternal incarceration on children's behavioral problems. *Demography* 51(3), 1041–1068.

Tables

TABLE 1: Summary statistics for 2010–2014 NC public school students

	(A)	(B)	(C)	(D)
	All NC students	Not linked to court records	Linked to court records	Exposed to active sentence
Panel A. Student characteristics				
Male	0.51	0.51	0.51	0.50
Age	11.76	11.76	11.75	11.69
Age at disposition			11.42	11.71
White	0.53	0.62	0.44	0.31
Black	0.27	0.20	0.33	0.51
Economically disadvantaged	0.51	0.41	0.62	0.84
Panel B. Student outcomes				
Math score	-0.04	0.11	-0.18	-0.47
English score	0.00	0.15	-0.15	-0.45
Days absent	7.31	6.72	7.97	9.66
Any suspension	0.08	0.06	0.11	0.17
Number of suspension days	0.76	0.56	0.98	1.80
Fighting incident	0.024	0.018	0.032	0.058
Panel C. Defendant/offense characteristics				
Male			0.64	0.83
Age at disposition			32.88	31.30
White			0.49	0.36
Black			0.36	0.57
Felony offense			0.16	0.53
Misdemeanor offense			0.23	0.33
Traffic offense			0.59	0.10
Clerk to decide offense			0.02	0.04
Months from arrest to disposition (median)			3.50	6.70
Guilty verdict			0.46	1.00
Active sentence			0.15	1.00
Minimum sentence in months (median)			1.67	2.00
Minimum sentence in months (mean)			10.41	10.97
# students	2,191,669	1,186,040	891,764	70,434
# math scores	3,903,718	1,763,568	1,834,605	179,412
# defendants			494,285	38,675

Notes: This table displays summary statistics on student characteristics (Panel A), student outcomes (Panel B), and characteristics of matched defendants and their offenses (Panel C). Column (A) includes all students who attended a North Carolina public school in 2010–2014. Column (B) includes the subset of these students who were not linked to a defendant in our criminal court records, and column (C) includes linked students. Columns (B)–(C) exclude students at schools that do not consistently report student addresses. Column (D) shows the subset of students in column (C) who experienced an active sentence in their household at any point in 2010–2014. See Appendices C.2.2, C.4, and C.5 for details on the data, merge, and sample definitions.

Math scores include scores on both end-of-grade 3–8 math exams and the end-of-course high school algebra exam. English scores include scores on both end-of-grade 3–8 reading exams and the end-of-course high school English exam. We standardize these scores to have a mean of zero and a standard deviation of one in the full population of test takers in each year. In Panel B, student outcomes are averaged over all years in 2010–2014. The defendant/offense characteristics in Panel C correspond to the most serious offense linked to each child, as described in Appendix C.5. See Appendix C.1 for details on variable definitions.

TABLE 2: OLS estimates of exposure to incarceration

Dependent variable	(A) No controls	(B) Demo- graphic controls	(C) County & year dummies	(D) No direct exposure	(E) Not in court records
Panel A. School exposure: Linear effect of 1 log point in HH incarcerations/student					
Math score	-0.205*** (0.009)	-0.104*** (0.007)	-0.104*** (0.006)	-0.102*** (0.006)	-0.098*** (0.006)
English score	-0.208*** (0.006)	-0.098*** (0.005)	-0.094*** (0.004)	-0.092*** (0.004)	-0.085*** (0.005)
N (math scores)	3,498,907	3,498,900	3,498,900	3,320,477	1,697,216
Panel B. Neighborhood exposure: Linear effect of 1 log point in incarcerations/resident					
Math score	-0.276*** (0.005)	-0.128*** (0.004)	-0.120*** (0.003)	-0.119*** (0.003)	-0.121*** (0.004)
English score	-0.278*** (0.005)	-0.127*** (0.003)	-0.117*** (0.003)	-0.116*** (0.003)	-0.113*** (0.003)
N (math scores)	3,271,929	3,271,922	3,271,922	3,101,129	1,548,345
Included fixed effects:					
Demographics		Y	Y	Y	Y
County & year			Y	Y	Y

Notes: This table shows a regression of student test scores on measures of school and neighborhood exposure to incarceration. We run the specification $Y_{igt} = \beta \log I_{gt} + \gamma_t + \gamma_{c(g)} + \mathbf{x}_i' \Phi + \epsilon_{igt}$. Our variable of interest is $\log I_{gt}$, which is a logged measure of incarceration exposure in school/neighborhood g and year t . In Panel A, I_{gt} is the number of active sentences linked to students who attended school g in the year of the defendant's disposition t . In Panel B, I_{gt} is the number of active sentences for criminal defendants who lived in census tract g in their disposition year t . Columns (B)–(E) include demographic controls, \mathbf{x}_i , which are fixed effects for gender, race, socioeconomic status, and birth year \times month. Columns (C)–(E) include fixed effects for year, γ_t , and county, $\gamma_{c(g)}$. The outcome variables, Y_{igt} , are math and English scores in standard deviation units, defined as in Table 1. The sample for columns (A)–(C) includes all North Carolina public school students with a math or English score in 2010–2014, but students in schools/neighborhoods with $I_{gt} = 0$ are omitted due to the log specification. Column (D) excludes students who were ever linked to a defendant with an active sentence in this time period. Column (E) excludes students who were ever linked to *any* defendant in our criminal court records. Parentheses contain standard errors clustered at the school (Panel A) and census tract (Panel B) level.

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

TABLE 3: Effects of county stringency on incarceration exposure and academic achievement

	(A)	(B)	(C)	(D)
Dependent variable	2010 mean (levels)	All students	Linked to court records	Not linked to court records
Panel A. Incarceration exposure				
$\Delta \log \#$ active sentences	1821	0.189*** (0.054)	0.189*** (0.054)	0.189*** (0.054)
$\Delta \#$ direct exposures / student	0.023	0.003** (0.001)	0.005** (0.002)	0.000
$\Delta \#$ classmate exposures / student	4.103	0.561*** (0.194)	0.569*** (0.205)	0.553*** (0.195)
N (# county/years)	100	400	400	400
Panel B. Academic performance				
Δ mean math score	-0.031	-0.024*** (0.008)	-0.027*** (0.008)	-0.027*** (0.007)
Δ mean English score	0.001	-0.014** (0.007)	-0.013 (0.008)	-0.015** (0.006)
N (# county/years)	100	400	400	400
# math scores	778,942	3,903,718	1,834,605	1,763,568

Notes: This table shows the impacts of changes in county stringency from 2010–2014, measured using a judge turnover strategy, on changes in student incarceration exposure and test scores. We follow Equation 2 and estimate the regression $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$. γ_t are year fixed effects. ΔS_{ct} corresponds to our main treatment variable, estimated as in equation 1 and discussed in Section 3.1, based on the change in local judicial stringency resulting from judicial arrivals/departures: $\Delta S_{ct} = \sum_j (\omega_{jct} - \omega_{jct-1}) \mu_{jct}^{-\{t-1,t\}}$. Column (B) includes all North Carolina public school students. Column (C) includes students who link to a defendant in the court records, and column (D) includes unlinked students (see columns (B)–(C) of Table 1). Across all specifications, we use a jackknife leave-out estimate of judicial stringency $\mu_{jct}^{-\{t-1,t\}}$. We scale ΔS_{ct} so that one unit represents one standard deviation of the distribution of county stringency. In Panel A, outcome variables include the log total number of active sentences in the county and the number of direct and classmate incarceration exposures per student; students are directly exposed when they experience an active sentence in their household, and they are indirectly exposed through a classmate if another student in their school and grade experiences a direct exposure. In Panel B, outcome variables are math and English scores in standard deviation units, defined as in Table 1. Parentheses contain standard errors clustered at the county level.

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

TABLE 4: County stringency balance tests

Dependent variable	2010 mean (levels)	Coefficient on Δ county stringency (SD units)			
		Benchmark specification		Fixed caseloads	
		Coef	(SE)	Coef	(SE)
Panel A. Criminal case totals					
Δ log # total cases	20,123	0.023	(0.019)	0.007	(0.011)
Δ log # felony cases	2,593	0.013	(0.027)	-0.020	(0.036)
Δ log # misdemeanor cases	9,308	0.026	(0.018)	-0.000	(0.014)
Δ log # traffic cases	7,579	0.024	(0.020)	0.031	(0.022)
Δ log # clerk to decide cases	642	0.142	(0.166)	0.114	(0.148)
<i>F</i> statistic: All coefficients zero		1.1		1.3	
Panel B. Employment					
Δ log # in labor force	168,882	-0.007	(0.005)	-0.005	(0.005)
Δ log # employed	152,170	-0.006	(0.006)	-0.003	(0.005)
Δ log # unemployed	16,712	-0.011	(0.007)	-0.006	(0.010)
<i>F</i> statistic: All coefficients zero		1.7		1.1	
Panel C. Defendant characteristics					
Δ proportion male	0.704	-0.001	(0.002)	0.004	(0.003)
Δ mean age	32.420	-0.051	(0.047)	-0.003	(0.068)
Δ proportion white	0.476	0.003	(0.006)	0.004	(0.007)
Δ proportion Black	0.433	-0.003	(0.007)	-0.003	(0.007)
Δ proportion w/ multiple offenses	0.432	-0.007*	(0.004)	-0.005	(0.004)
Δ proportion w/ prior offense	0.505	-0.001	(0.005)	-0.001	(0.005)
Δ proportion w/ prior active sentence	0.072	0.005	(0.005)	0.003	(0.007)
<i>F</i> statistic: All coefficients zero		1.0		1.2	
Panel D. Student characteristics					
Δ proportion male	0.507	-0.000	(0.001)	0.000	(0.001)
Δ mean age	11.764	0.005	(0.008)	-0.001	(0.011)
Δ proportion white	0.541	-0.002	(0.001)	-0.002*	(0.001)
Δ proportion Black	0.270	0.003**	(0.001)	0.003	(0.002)
Δ proportion economically disadvantaged	0.492	0.001	(0.002)	0.002	(0.002)
Δ mean predicted Math score	-0.059	-0.001	(0.001)	-0.001	(0.001)
Δ mean predicted English score	-0.005	-0.001	(0.001)	-0.001	(0.001)
<i>F</i> statistic: All coefficients zero		1.5		1.0	
<i>N</i> (# county/years)	100	400		400	
# criminal cases	680,192	3,153,479		3,153,479	
# students \times years	1,457,836	7,422,413		7,422,413	

Notes: This table examines how changes in our county-level measure of judge stringency are related to changes in active sentences, criminal case totals, defendant characteristics, and student characteristics. We estimate Equation 2, $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$. ΔS_{ct} is the change in county stringency using both actual caseloads (column (B)) and fixed caseloads (column (C)), and γ_t are year fixed effects. Our outcomes, ΔY_{ct} , measure changes in four sets of variables. In Panel A, the outcome variable is changes in county-level log criminal case totals. Panel B investigates changes in county-level labor force and employment totals from the Bureau of Labor Statistics' LAUS database. Panel C shows changes in the average characteristics of criminal defendants. Panel D investigates changes in the mean characteristics of all North Carolina public school students (column (A) in Table 1). In the last two rows of Panel D, the dependent variables are predicted values from regressions of Math/English scores on age and dummies for gender, race/ethnicity, economic disadvantage, limited English proficiency, and disability codes. For all panels, we present *F*-statistics from a test of joint significance of all coefficients. Parentheses contain standard errors clustered at the county level.

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

TABLE 5: Isolating specific sources of variation in county stringency

	(A)	(B)	(C)	(D)	(E)	(F)
Dependent variable	Bench- mark	Fixed caseloads	Only departing judges	Judicial elections	Close elections	Retire- ment elections
$\Delta \log \#$ active sentences	0.189*** (0.054)	0.137** (0.061)	0.202** (0.084)	0.188*** (0.051)	0.182*** (0.067)	0.194*** (0.057)
Δ mean math score	-0.024*** (0.008)	-0.036** (0.018)	-0.047* (0.026)	-0.034** (0.014)	-0.043** (0.019)	-0.029** (0.013)
Δ mean English score	-0.014** (0.007)	-0.020* (0.010)	-0.027 (0.017)	-0.018** (0.007)	-0.019* (0.010)	-0.024*** (0.009)
N (# county/years)	400	400	400	400	400	400

Notes: This table show how different sources of variation in county stringency affect county-level active sentences and student test scores. Column (A) replicates our benchmark results from column (B) of Table 3, which use our actual caseloads measure of ΔS_{ct} . For our actual caseloads measure, the weights in equation (1), ω_{jct} , are equal to the proportion of cases in county c and year t that were heard by judge j . Column (B) uses our fixed caseload measure of ΔS_{ct} ; in this measure, the weights, ω_{jct} , are equal to zero if judge j is not active in county c in year t , and otherwise equal to judge j 's average fraction of cases heard in county c across all active years in our sample. Column (C) also uses our fixed caseloads measure, but we use only variation from departing judges, i.e., from judges who leave a county and do not return during the period of our data. Columns (D)–(F) use variation in fixed caseloads driven only by judges winning or losing elections. We identify judges who enter or exit due to elections by linking judges in the court data to biannual election results using judge initials, names, and districts. Column (D) uses variation from all elections. Column (E) uses variation only from close elections, defined as elections in the winning judge's share was below 55 percent. Column (F) uses only elections that occurred because the incumbent judge retired. We report estimates of β from equation (2) using each measure of ΔS_{ct} . We use year-to-year changes in three different county-level outcomes as dependent variables: log active sentences, mean math scores, and mean English scores. Parentheses contain standard errors clustered at the county level.

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

TABLE 6: Direct impacts of a household incarceration on student outcomes

Dependent variable	(A)	(B)	(C)	Judge stringency estimate	
	Mean	OLS coef. on active sentence	Event study estimate	Reduced form (SD units)	IV for active sentence
Panel A. First stage					
Active sentence	0.137			0.034*** (0.003)	
Panel B. Misbehavior and Attendance					
Any suspension	0.158	0.009*** (0.003)	0.013*** (0.002)	0.003** (0.001)	0.099** (0.043)
Number of suspension days	1.675	0.153*** (0.055)	0.252*** (0.036)	0.021 (0.022)	0.642 (0.666)
Fighting incident	0.048	0.003** (0.002)	0.002* (0.001)	0.001* (0.001)	0.042* (0.024)
Days absent	9.312	0.412*** (0.094)	0.456*** (0.059)	0.079* (0.045)	2.372* (1.396)
Panel C. Academic performance					
Math score	-0.359	-0.034*** (0.010)	-0.015*** (0.005)	-0.006 (0.006)	-0.164 (0.169)
English score	-0.308	-0.014 (0.011)	-0.012** (0.005)	-0.012** (0.006)	-0.348** (0.164)
Judge stringency sample	Y	Y		Y	Y
Event study sample			Y		
Years, t , relative to case disposition, $\tau(i)$	$t \geq \tau(i)$	$t \geq \tau(i)$	All t	$t \geq \tau(i)$	$t \geq \tau(i)$
N (math scores)	296,242	296,242	699,084	296,242	296,242
# students	118,416	118,416	128,829	118,416	118,416
F -statistics:					
First stage				103.0	
Defendant balance test				0.9	
Student balance test				1.1	

Notes: This table presents estimates of the direct impacts of a household incarceration using our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our direct judge stringency sample (column (B) in Appendix Table A5). Column (C) includes students in our direct event study sample (column (A) in Appendix Table A5). See Appendix C.5 for details on our samples. Each row corresponds to a separate regression using the dependent variable listed in the first column. Regressions are at the student \times year level.

Column (A) displays means of each outcome variable measured in calendar years, t , in or after the disposition year of the defendant's case, $\tau(i)$. Column (B) shows OLS estimates of β from Equation 3 using outcomes measured in years $t \geq \tau(i)$, and column (C) displays estimates of δ from the event study specification 4 using all years t . Column (D) shows reduced-form judge stringency estimates of γ from Equation 5 for years $t \geq \tau(i)$, with estimates normalized to represent a one standard deviation increase in judge stringency. Column (E) shows estimates of β from the IV specification based on Equations 3 and 5 using years $t \geq \tau(i)$. Regressions in columns (B), (D), and (E) include court \times year \times offense class dummies and defendant and student characteristics (age at disposition, gender, race dummies, student socioeconomic status, and missing values of each covariate). Parentheses contain standard errors clustered at the defendant (column (C)) and judge (columns (B), (D), and (E)) levels. The bottom of the table reports F -statistics from the first stage and joint significance tests for defendant and student characteristics for the specification in column (D); see Appendix Table B1 for details on the balance tests.

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

TABLE 7: Indirect impacts of a household incarceration on classmates

Dependent variable	(A)	(B)	(C)	Judge stringency estimate	
	Mean	OLS coef. on active sentence	Event study estimate	Reduced form (SD units)	IV for active sentence
Panel A. First stage					
Active sentence	0.133			0.031*** (0.002)	
Panel B. Academic performance					
Math score	-0.117	-0.006* (0.003)	-0.004** (0.002)	-0.001 (0.002)	-0.017 (0.062)
English score	-0.048	-0.008*** (0.003)	-0.003** (0.001)	-0.001 (0.002)	-0.047 (0.053)
Judge stringency sample	Y	Y		Y	Y
Event study sample			Y		
Years, t , relative to case disposition, $\tau(i)$	$t \geq \tau(i)$	$t \geq \tau(i)$	All t	$t \geq \tau(i)$	$t \geq \tau(i)$
N (math scores)	32,120,520	32,120,520	103,609,169	32,120,520	32,120,520
# students	1,454,577	1,454,577	1,470,600	1,454,577	1,454,577
<i>F</i> -statistics:					
First stage				189.6	
Defendant balance test				1.3	
Student balance test				1.2	

Notes: This table presents estimates of the indirect impacts of a household incarceration on classmates using our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our indirect judge stringency sample (column (E) in Appendix Table A5). Column (C) includes students in our indirect event study sample (column (D) in Appendix Table A5). These samples include children who attended the same school and grade as children in our direct exposure samples in the year of the defendant’s disposition (see Appendix C.5 for details). Each row corresponds to a separate regression using the dependent variable listed in the first column. Regressions are at the student \times directly impacted child \times year level; thus student \times year observations are repeated for each directly impacted child in their school/grade.

Column (A) displays means of each outcome variable measured in calendar years, t , in or after the disposition year of the defendant’s case, $\tau(i)$. Column (B) shows OLS estimates of β from Equation 7 using outcomes measured in years $t \geq \tau(i)$. Column (C) displays estimates of δ from the event study specification 6 using all years t . Column (D) shows reduced-form judge stringency estimates of γ from Equation 8 for years $t \geq \tau(i)$, with estimates normalized to represent a one standard deviation increase in judge stringency. Column (E) shows estimates of β from the IV specification based on Equations 7 and 8 using years $t \geq \tau(i)$. Regressions in columns (B), (D), and (E) include court \times year \times offense class dummies and defendant and student characteristics (age at disposition, gender, race dummies, student socioeconomic status, and missing values of each covariate). Parentheses contain standard errors clustered at the school level in columns (B)–(E). The bottom of the table reports *F*-statistics from the first stage and joint significance tests for defendant and student characteristics for the specification in column (D); see Appendix Table B1 for details on the balance tests.

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

Figures

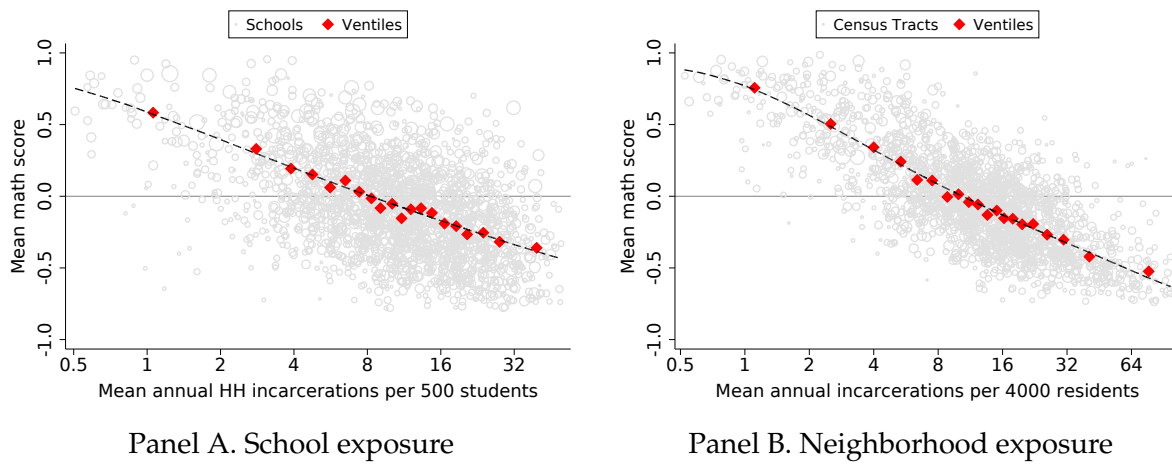
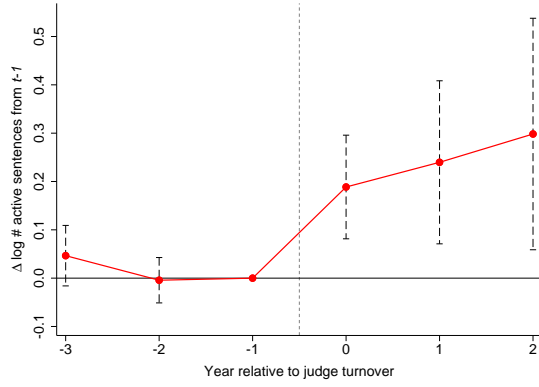
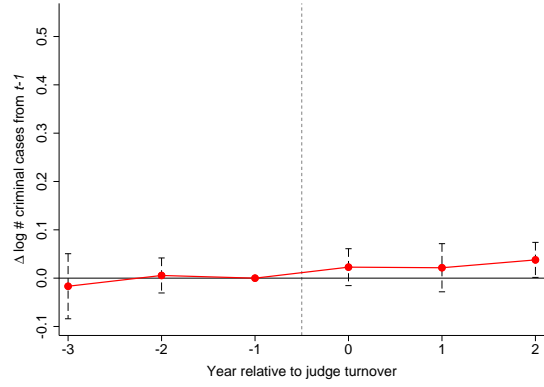


FIGURE 1: Math scores by school and neighborhood exposure to incarceration (log scale)

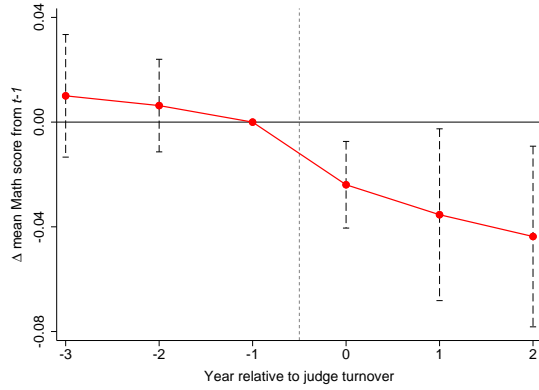
Notes: This figure plots the relationship between student math scores and exposure to incarceration at the school (Panel A) and neighborhood (Panel B) levels. In Panel A, gray circles represent schools, and the x -axis shows the average annual number of active sentences for defendants linked to students in each school (per 500 students—roughly the median school size). In Panel B, gray circles represent census tracts, and the x -axis shows the average annual number of active sentences for individuals with an address in the tract (per 4,000 residents—roughly the median tract size). In both panels, the y -axis depicts students' scores on both end-of-grade 3–8 math exams and the end-of-course high school algebra exam. We standardize these scores to have a mean of zero and a standard deviation of one in the full population of test takers in each year. The sample includes active sentences in 2010–2014 and all North Carolina public school students with math scores in those years. Both figures are trimmed at the 99th and 1st percentiles and use log scales on the x -axis and exclude observations below 0.5 mean annual incarcerations per 500 students/4,000 residents.



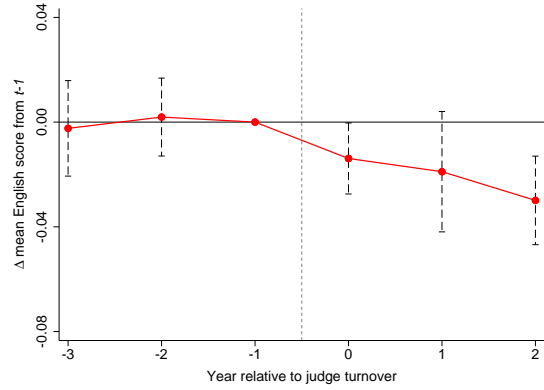
Panel A. $\Delta \log \#$ active sentences



Panel B. $\Delta \log \#$ criminal cases



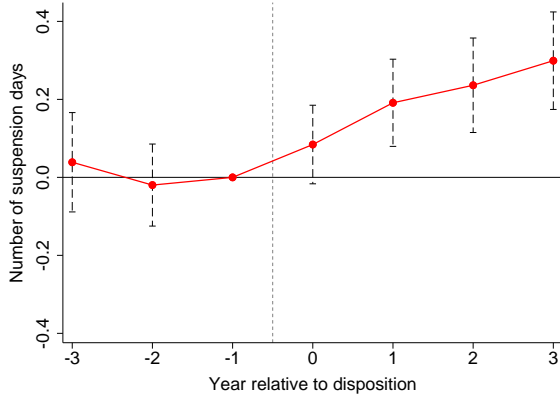
Panel C. Δ mean math score



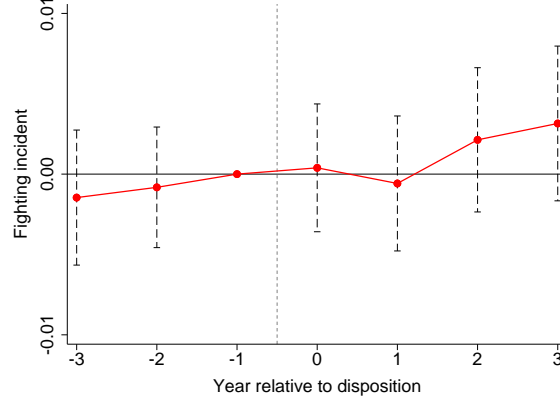
Panel D. Δ mean English score

FIGURE 2: Leading and lagged effects of a change in county stringency between years $t - 1$ and t

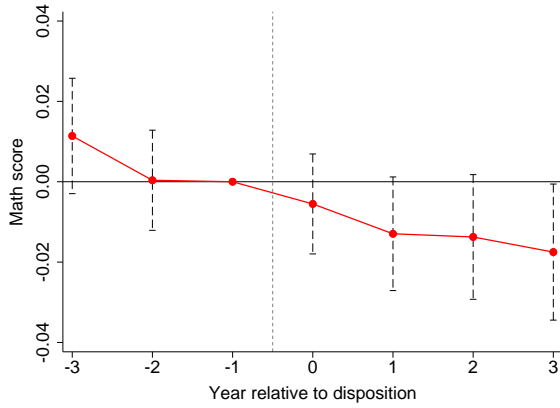
Notes: This figure reports estimates from a modified version of Equation 2 that examines leading and lagged effects of a change in county stringency. We run the specification $\Delta Y_{c,t+k} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$, where the dependent variable, $\Delta Y_{c,t+k}$, measures the cumulative change in the outcome in county c between year $t + k$ and year $t - 1$. We consider values of k from -3 to 2 , as indicated by the x -axis of each panel. We use four outcome variables: log number of active sentences (Panel A), log number of criminal cases (Panel B), mean math scores (Panel C), and mean English scores (Panel D). In all regressions, the variable of interest, ΔS_{ct} , is defined as the change in county stringency between year t and $t - 1$. We compute a separate version of ΔS_{ct} for each regression in which the underlying individual judge stringency measures, $\mu_{jct}^{-\{t-1, \dots, t+k\}}$, are computed leaving out all observations between years $t + k$ and $t - 1$. Vertical dashes lines are 95 percent confidence intervals with standard errors clustered at the county level.



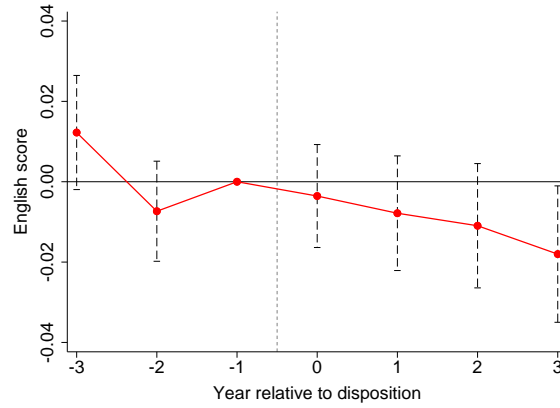
Panel A. Number of suspension days



Panel B. Fighting incident



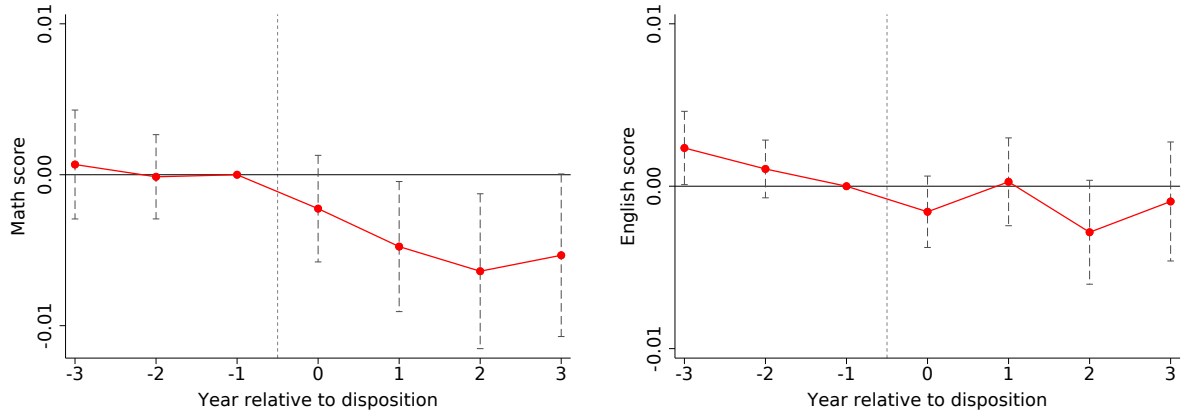
Panel C. Math score



Panel D. English score

FIGURE 3: Direct impacts of a household incarceration: Event study

Notes: This figure shows event studies of the direct impacts of a household incarceration on student outcomes. We use our direct exposure event study sample (column (A) in Appendix Table A5) and estimate the following regression: $y_{it} = \sum_{l=-8}^9 \delta_l \mathbb{1}\{t - \tau(i) = l\} \times \text{Active Sentence}_i + \eta_i + \lambda_{og\tau(i)t} + \varepsilon_{it}$. These specifications show outcomes y_{it} for student i in calendar year t who has a household member that receives a disposition in year $\tau(i)$. Active Sentence_i is an indicator equal to one if the defendant linked to student i received an active sentence. We interact this term with dummies for years l relative to the defendant's disposition, omitting the $l = -1$ term. η_i is an individual fixed effect. $\lambda_{og\tau(i)t}$ is an offense class \times academic cohort \times disposition year \times calendar year fixed effect. The graphs plot the δ_l coefficients from $l = -3$ to 3. The outcome variable for each regression, y_{it} , is listed in the panel title; see Appendix C.1 for details on variable definitions. Dashed lines are 95 percent confidence intervals using standard errors clustered at the defendant level.



Panel A. Classmates' math scores

Panel B. Classmates' English scores

FIGURE 4: Indirect impacts of a household incarceration: Event study and timing of judge stringency effects

Notes: This figure shows event study and judge stringency estimates of the indirect impacts of a household incarceration on academic outcomes. Panels A–B include the classmates of students in our direct event study sample, where classmates are defined as children who attended the same school and grade as children in our direct exposure samples in the year of the defendant's disposition (see Appendix C.5). Panels A–B present event study estimates using a dynamic version of equation 6: $y_{kit} = \sum_{l=-8}^9 \delta_l \cdot \mathbb{1}\{t - \tau(i) = l\} \times \text{Active Sentence}_i + \eta_{ki} + \lambda_{og\tau(i)t} + \varepsilon_{kit}$. These specifications show outcomes y_{kit} for student k in calendar year t whose classmate/neighborhood peer i is linked to a defendant that receives a disposition in year $\tau(i)$. Active Sentence_i is an indicator equal to one if the defendant linked to peer i received an active sentence. We interact this variable with indicators for years relative to disposition $l = -8$ to 9 , omitting the $l = -1$ interaction. η_{ik} is a student \times year fixed effect, and $\lambda_{og\tau(k)t}$ is an offense class \times academic cohort \times disposition year \times calendar year fixed effect. The outcome variables are math and English scores in standard deviation units. The graphs plot the δ_l and γ_l coefficients from $l = -3$ to 3 . Dashed lines are 95 percent confidence intervals using standard errors clustered at the school level.

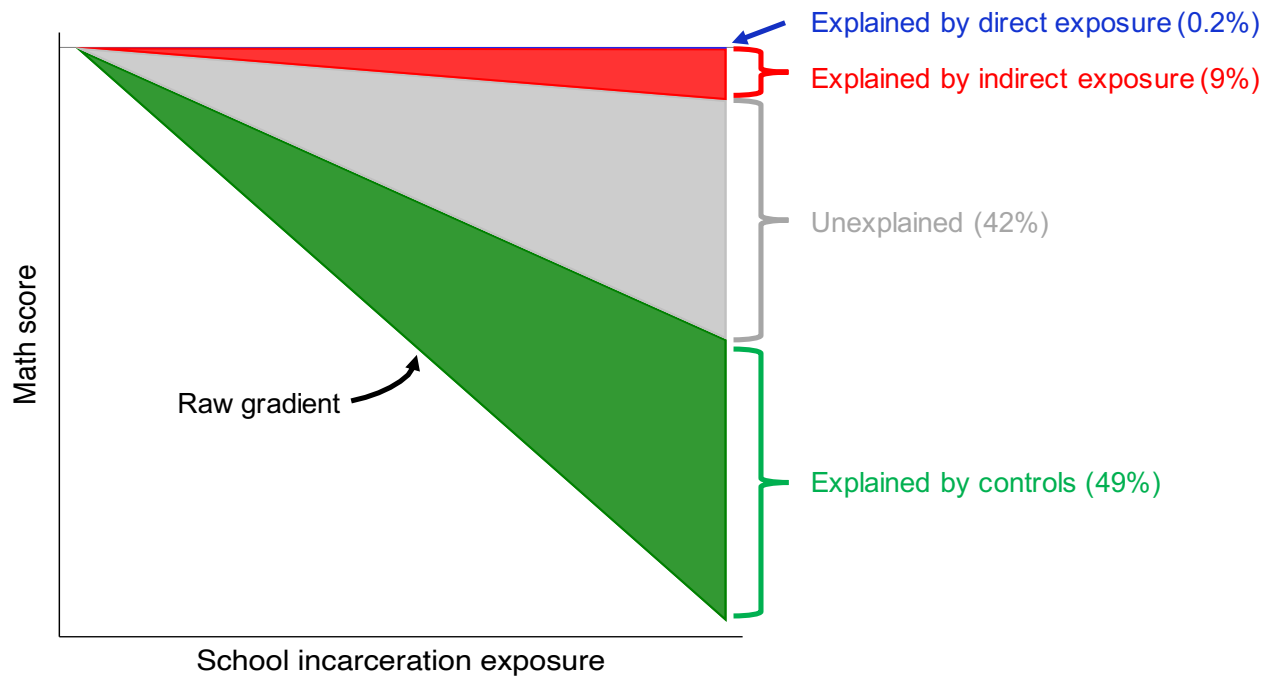


FIGURE 5: Decomposition of the gradient between math scores and school incarceration exposure

Notes: This figure shows the proportion of the raw gradient between math scores and school incarceration exposure that can be explained by our direct and indirect event study estimates. The raw gradient is the linear relationship between student math scores and log household incarcerations per student from column (A) of Table 2 (-0.205). The green region shows that demographic, county, and year controls can explain 49 percent of this gradient; this comes from the estimate in column (C) of Table 2 (-0.104), which is 51 percent of the raw gradient. The blue region shows that our event study estimate of the direct impact of a household incarceration can explain 0.2 percent of the raw gradient. The red region shows that our event study estimate of the indirect impacts of a household incarceration can explain 9 percent of the raw gradient. We compute the values of the blue and red regions by translating the event study estimates and the raw gradient into effects of one additional household incarceration at the mean values of incarceration exposure; see Appendix Table A9 for details on these calculations. The gray region shows the remaining unexplained component (42 percent of the raw gradient).

A Appendix Tables and Figures

TABLE A1: Sources of variation in county stringency

Dependent variable	(A)	(B)	(C)	(D)	(E)
	Actual caseloads	Fixed caseloads	Sources of variation in fixed caseloads		
			Judicial elections	Mid-term leaves	County rotations
Panel A. Variance decomposition for change in county stringency (ΔS_{ct})					
SD of change in county stringency (ΔS_{ct})	0.0180	0.0140	0.0077	0.0064	0.0032
Variance of ΔS_{ct}	32.5E-5	19.6E-5	5.9E-5	4.1E-5	1.0E-5
% of total variance in ΔS_{ct}	100%	60%	18%	13%	3%
Panel B. Main effects (scaled to 1 SD of county stringency in each measure)					
$\Delta \log \#$ active sentences	0.189*** (0.054)	0.137** (0.061)	0.209** (0.081)	0.172 (0.145)	0.206* (0.115)
Δ mean math score	-0.024*** (0.008)	-0.036** (0.018)	-0.035* (0.018)	-0.100** (0.044)	0.042 (0.052)
Δ mean English score	-0.014** (0.007)	-0.020* (0.010)	-0.027** (0.013)	-0.046** (0.023)	0.033 (0.047)
<i>N</i> (# county/years)	400	400	400	400	400
# math scores	3,903,718	3,903,718	3,903,718	3,903,718	3,903,718
Panel C. Balance tests (<i>F</i>-statistics)					
Criminal case totals	0.3	1.5	2.3	0.5	1.2
Defendant characteristics	1.0	1.2	1.6	2.2	2.2
Student characteristics	1.5	1.0	0.6	0.7	1.5
Prior year test scores	0.5	2.0	0.9	1.0	1.3

Notes: This table decomposes the change in county stringency, ΔS_{ct} , into its underlying components and estimates the effects of each component on student test scores. We begin in column (A) with the actual caseload measure, which uses variation in both whether or not a judge serves in a county in a given year as well as their number of cases that county/year. Column (B) reports the fixed caseload measure, which restricts to only the variation in whether or not a judge serves in a county (by using, in each year, the judge's mean caseload across all years). Columns (C)–(E) decompose the fixed caseload measure in column (B) into three components. Column (C) computes ΔS_{ct} using only judges who began or stopped working in *any* county in our data due to an election win or loss. We compute this measure by linking judges in the court data to biannual election results using judge initials, names, and districts. Column (D) uses variation from mid-term leaves, which we define as all reasons that judges began or stopped working in *any* county in our data other than elections. Last, column (E) uses variation from county rotations, which are judges who began or stopped working in a *given* county but still appear in our data in other counties.

Panel A reports the standard deviation of the *change* in county stringency, ΔS_{ct} (first row). This panel also reports the percentage of the total variance in ΔS_{ct} explained by each component, which is equal to the variance of ΔS_{ct} in each column divided by the variance of ΔS_{ct} in column (A). Panel B shows the impacts of each of these underlying forms of variation on active sentences and test scores; these are analogous to the coefficients in Tables 3 and 4, but we use ΔS_{ct} computed from the source of variation highlighted in each column. We normalize each ΔS_{ct} measure so that one unit corresponds to one standard deviation in county stringency for that measure. Parentheses contain standard errors clustered at the county level. Panel C reports *F*-statistics from balance tests for the joint significance of changes in criminal case totals, defendant characteristics, student characteristics, and prior year test scores. These balance tests are analogous to those in Tables 4 and A3, but again use ΔS_{ct} computed from the source of variation for each column.

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

TABLE A2: Counties with large changes in mean stringency and effects on outcomes

Panel A. Counties with large year-to-year changes in mean stringency

County	(A) # annual cases	(B) Year of turnover event	(C) Change in county stringency	(D) Event
<i>Increases in stringency (90th percentile change or higher)</i>				
Buncombe	7,424	2011	0.036	EC & WS appointed by Gov. Purdue to replace SHB & STB
Swain	562	2012	0.033	RTW (53.9% of vote) won election to replace SJB
Graham	176	2013	0.054	Rotation of JUD (superior court) & RDH (district court)
Carteret	3,038	2013	0.042	CR (52.8%) defeated CS; WM (71.0%) won election to replace JW
Pamlico	363	2013	0.035	CR (52.8% of vote) defeated CS
Lenoir	4,077	2014	0.154	EJ (52.6% of vote) defeated LC
Cabarrus	8,720	2014	0.033	CEW appointed by Gov. McCrory to replace MBM
Median			0.036	
<i>Decreases in stringency (10th percentile change or lower)</i>				
Northampton	794	2012	-0.023	TLJ (60.2% of vote) defeated AWK
Clay	144	2013	-0.050	Rotation of RWK (district court)
Durham	8,557	2013	-0.017	Rotation of OFH, EMB, & MO (superior court)
Pitt	9,095	2014	-0.112	LFT appointed by Gov. McCrory to replace CMV
Gaston	6,687	2014	-0.016	Rotation HBL & LB (superior court)
Gates	229	2014	-0.016	Rotation of ELB (district court)
Median			-0.020	

Panel B. Impacts on sentencing, criminality, and children's academic outcomes

Covariate	(A)	(B)	(C)	(D)
	Dependent variable			
	Log # active sentences	Log # criminal cases	Average math score	Average English score
1(Increase in stringency) × Post	0.212*** (0.062)	0.023 (0.015)	-0.051** (0.023)	-0.030*** (0.011)
1(Decrease in stringency) × Post	-0.086 (0.159)	0.007 (0.045)	0.026 (0.026)	0.016 (0.021)
N (# county/years)	500	500	1,200	1,200

Notes: This table identifies counties that experienced large changes in mean judge stringency and shows the effects of these events on county-level outcomes. Panel A shows counties that experienced judge turnover events that significantly increased or decreased county average stringency. To identify these events, we compute the year-to-year changes in county stringency for all counties and years in our data. We define large increases in stringency as changes at the 90th percentile of this distribution or above that persisted in magnitude through 2014. We also define large decreases in stringency as changes at the 10th percentile of this distribution or below that persisted in magnitude through 2014. Column (A) shows the number of cases heard in each county that experienced one of these events in the year of the event, shown in column (B). Column (C) shows the county's year-to-year change in stringency in that year. Column (D) describes the event(s) that were most responsible for the change in stringency; we identified these events by searching the web for judicial election results and news stories on judge appointments. "Rotations" are events in which a judge continued to hold their judicial appointment but heard 20+ cases in the county in only one of the two years.

Panel B displays the impacts of the judge turnover events on county-level active sentences, criminal cases, and the average math and English scores of children in those counties. We regress the dependent variable listed in the column header on year dummies, county dummies, 1(Increase in stringency) × Post, and 1(Decrease in stringency) × Post. 1(·) is an indicator for a county experiencing each type of event, and Post is an indicator for years after the event. We display the coefficients on these two interaction variables. Regressions are at the county/year level. Columns (A)–(B) include all years in which we observe court data (2009–2014). Columns (C) include all years in which we observe children's test scores (2006–2017). Parentheses contain standard errors clustered at the county level.

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

TABLE A3: Leading and lagged effects of a change in county stringency between years $t - 1$ and t

Dependent variable	(A)	(B)	(C)	(D)	(E)	
	2010 mean (levels)	Change in outcome between this year and $t - 1$				
		$t - 3$	$t - 2$	t	$t + 1$	$t + 2$
Panel A. Active sentences						
$\Delta \log \#$ active sentences	1907	0.047 (0.032)	-0.004 (0.024)	0.189*** (0.054)	0.240*** (0.085)	0.298*** (0.121)
Panel B. Academic performance						
Δ mean math score	-0.031	0.010 (0.012)	0.006 (0.009)	-0.024*** (0.008)	-0.035** (0.017)	-0.044** (0.017)
Δ mean English score	0.001	-0.002 (0.009)	0.002 (0.008)	-0.014** (0.007)	-0.019 (0.012)	-0.030*** (0.009)
Panel C. Criminal case totals						
$\Delta \log \#$ total cases	20534	-0.017 (0.034)	0.006 (0.018)	0.023 (0.019)	0.022 (0.025)	0.038** (0.018)
$\Delta \log \#$ felony cases	2643	-0.040 (0.041)	-0.047 (0.040)	0.013 (0.027)	0.004 (0.016)	0.016 (0.016)
$\Delta \log \#$ misdemeanor cases	9420	-0.020 (0.041)	0.021 (0.024)	0.026 (0.018)	0.025 (0.024)	0.016 (0.016)
$\Delta \log \#$ traffic cases	7803	0.041 (0.042)	0.020 (0.026)	0.024 (0.020)	0.013 (0.035)	0.083*** (0.028)
$\Delta \log \#$ clerk to decide cases	667	-0.164 (0.150)	-0.031 (0.094)	0.142 (0.166)	0.244 (0.250)	0.229 (0.272)
Panel D. Employment						
$\Delta \log \#$ in labor force	168882	-0.006 (0.006)	-0.005 (0.005)	-0.007 (0.005)	-0.001 (0.012)	-0.012 (0.017)
$\Delta \log \#$ employed	152170	-0.011* (0.006)	-0.006 (0.005)	-0.006 (0.006)	0.001 (0.011)	-0.009 (0.017)
$\Delta \log \#$ unemployed	16712	0.007 (0.013)	0.011 (0.009)	-0.011 (0.007)	-0.025 (0.024)	-0.030 (0.023)
N (# county / years)	100	400	400	400	400	400

Notes: This table reports estimates from a modified version of Equation 2 that examines leading and lagged effects of a change in county stringency. We run the specification $\Delta Y_{c,t+k} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$, where the dependent variable, $\Delta Y_{c,t+k}$, measures the cumulative change in the outcome in county c between year $t + k$ and year $t - 1$. We consider values of k from -3 to 2 , as indicated by the headers of columns (B) through (E). We use four types of outcome variables: log number of active sentences (Panel A), mean student test scores (Panel B), log criminal case totals (Panel C), and log employment totals (Panel D). In all columns, the variable of interest, ΔS_{ct} , is defined as the change in county stringency between year t and $t - 1$. We compute a separate version of ΔS_{ct} for each column in which the underlying individual judge stringency measures, $\mu_{jct}^{-\{t-1, \dots, t+k\}}$, are computed leaving out all observations between years $t + k$ and $t - 1$. Parentheses contain standard errors clustered at the county level.

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

TABLE A4: Heterogeneity in county stringency effects by exposure to criminal cases

	(A)	(B)	(C)	(D)	(E)
	Coefficient on Δ county stringency (SD units)				
Panel A. Heterogeneity by school exposure to criminal cases					
		Quartiles of criminal cases per student in 2010			
Dependent variable	All schools	Bottom qtile	Q2	Q3	Top qtile
Δ log HH incarcerations per student	0.126*** (0.038)	0.121 (0.141)	0.139* (0.079)	0.075 (0.051)	0.296** (0.109)
Δ HH incarcerations per 500 students	1.031** (0.485)	0.562 (1.317)	1.481 (1.124)	0.187 (0.456)	4.657*** (1.475)
Δ mean math score	-0.028*** (0.008)	-0.025 (0.048)	-0.015 (0.009)	-0.027** (0.009)	-0.059** (0.024)
Δ mean English score	-0.016** (0.006)	-0.005 (0.038)	-0.022** (0.010)	-0.005 (0.004)	-0.028 (0.025)
N (# school/years)	9,112	2,324	2,236	2,288	2,264
Panel B. Heterogeneity by neighborhood exposure to criminal cases					
		Quartiles of criminal cases per resident in 2010			
Dependent variable	All tracts	Bottom qtile	Q2	Q3	Top qtile
Δ log incarcerations per resident	0.142*** (0.031)	0.172** (0.070)	0.132*** (0.033)	0.138*** (0.036)	0.095** (0.040)
Δ incarcerations per 4,000 residents	2.332*** (0.546)	1.254*** (0.279)	1.897*** (0.405)	2.321*** (0.689)	3.622** (1.742)
Δ mean math score	-0.026*** (0.008)	-0.019** (0.009)	-0.026** (0.011)	-0.024** (0.011)	-0.028*** (0.009)
Δ mean English score	-0.016*** (0.006)	-0.020** (0.010)	-0.012* (0.007)	-0.016 (0.011)	-0.018* (0.010)
N (# tract/years)	8,596	2,152	2,148	2,148	2,148

Notes: This table examines heterogeneity in the effects of county stringency on student test scores by exposure to criminal cases. The table displays estimates of β from Equation 2, but outcomes are defined at the school (Panel A) or census tract (Panel B) level rather than at the county level. The outcome variables are changes in incarceration exposure per student/resident (in logs and levels) and changes in math/English scores. The sample for column (A) includes all schools/tracts that have students with test scores in each year in 2010–2014. In columns (B)–(E), we divide schools/tracts into quartiles based on the number of criminal cases per student/resident in 2010. Parentheses contain standard errors clustered at the county level.

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

TABLE A5: Summary statistics for direct and indirect analysis samples

	(A)	(B)	(C)	(D)	(E)
	Direct exposure samples			Indirect exposure samples (classmates)	
	Event study	Judge stringency	Judge compliers	Event study	Judge stringency
Panel A. Student characteristics					
Male	0.50	0.50	0.51	0.51	0.51
Age	12.15	12.15	12.06	12.03	12.04
Age at disposition	12.89	12.85	12.70		
White	0.43	0.42	0.30	0.54	0.54
Black	0.39	0.40	0.53	0.26	0.26
Economically disadvantaged	0.72	0.72	0.75	0.49	0.49
Panel B. Student outcomes					
Math score	-0.32	-0.32		0.01	0.01
English score	-0.28	-0.29		0.03	0.03
Days absent	8.79	8.80		7.10	7.10
Any suspension	0.14	0.14		0.08	0.08
Number of suspension days	1.56	1.55		0.86	0.86
Fighting incident	0.045	0.045		0.025	0.025
Panel C. Defendant/offense characteristics					
Male	0.73	0.70	0.74		
Age at disposition	32.72	32.75	31.78		
White	0.48	0.47	0.35		
Black	0.42	0.43	0.60		
Felony offense	0.30	0.28	0.32		
Misdemeanor offense	0.38	0.35	0.53		
Traffic offense	0.27	0.33	0.10		
Clerk to decide offense	0.05	0.04	0.04		
Months from arrest to disposition (median)	6.37	6.23			
Guilty verdict	1.00	0.79			
Active sentence	0.29	0.14			
Minimum sentence in months (median)	1.50	3.03			
Minimum sentence in months (mean)	10.01	13.82			
# students	128,829	118,416		1,470,600	1,454,577
# math scores	699,084	643,595		6,951,743	6,905,245
# defendants	86,854	79,765			

Notes: This table displays summary statistics on our direct and indirect analysis samples for Sections 4–5. It shows student characteristics (Panel A), student outcomes (Panel B), and characteristics of matched defendants and their offenses (Panel C). Column (A) shows our event study sample for the direct impacts of a household incarceration, which includes students whose linked defendant was convicted of their most serious offense. Column (B) shows our judge stringency sample for direct impacts, which includes students whose linked defendant faced a judge for which we can compute a stringency measure. Column (C) displays mean complier characteristics for the judge stringency sample in column (B) using the methodology of [Frandsen, Lefgren, and Leslie \(2019\)](#); we estimate our judge stringency IV specification (Equations 3 and 5), where the dependent variable is the student/defendant characteristic interacted with an indicator for an active sentence, and display the resulting IV coefficient in column (C). Columns (D)–(E) include students who were in the same school and grade as the children in columns (A)–(B) in the year of the defendant’s disposition; these are our samples for examining the indirect impacts of a household incarceration. See Appendices C.2.2, C.4, and C.5 for details on the data, merge, and sample definitions.

Math scores include scores on both end-of-grade 3–8 math exams and the end-of-course high school algebra exam. English scores include scores on both end-of-grade 3–8 reading exams and the end-of-course high school English exam. We standardize these scores to have a mean of zero and a standard deviation of one in the full population of test takers in each year. In Panel B, student outcomes are averaged over all years in 2006–2017. The defendant/offense characteristics in Panel C correspond to the most serious offense linked to each child, as described in Appendix C.5. See Appendix C.1 for details on variable definitions.

TABLE A6: Event study robustness tests

Dependent variable	(A) No offense controls	(B) Offense class at arrest	(C) Offense class at conviction	(D) Convicted 4-digit offense	(E) Controls for prior points
Panel A. Direct effects of a household incarceration					
Math score	-0.0143*** (0.0051)	-0.0118** (0.0055)	-0.0154*** (0.0054)	-0.0134** (0.0061)	-0.0107* (0.0062)
English score	-0.0149*** (0.0048)	-0.0106** (0.0052)	-0.0122** (0.0052)	-0.0114* (0.0059)	-0.0076 (0.0060)
Days absent	0.5157*** (0.0542)	0.4265*** (0.0590)	0.4556*** (0.0586)	0.3217*** (0.0648)	0.3044*** (0.0661)
Any suspension	0.0163*** (0.0018)	0.0129*** (0.0019)	0.0133*** (0.0019)	0.0089*** (0.0021)	0.0072*** (0.0021)
Number of suspension days	0.2981*** (0.0338)	0.2301*** (0.0367)	0.2523*** (0.0363)	0.2009*** (0.0398)	0.1438*** (0.0407)
Fighting incident	0.0027** (0.0011)	0.0021* (0.0011)	0.0020* (0.0011)	0.0011 (0.0013)	0.0007 (0.0013)
Neighborhood SES percentile	0.0008 (0.0016)	0.0017 (0.0017)	0.0028* (0.0017)	0.0044** (0.0019)	0.0045** (0.0019)
Student has criminal charge	-0.0026* (0.0014)	-0.0022 (0.0015)	-0.0019 (0.0015)	-0.0023 (0.0017)	-0.0020 (0.0017)
N (math scores)	696,842	696,071	696,289	655,070	655,070
Panel B. Indirect effects of a household incarceration on classmates					
Math score	-0.0062*** (0.0019)	-0.0052*** (0.0019)	-0.0041** (0.0019)	-0.0036* (0.0019)	-0.0032* (0.0019)
English score	-0.0038*** (0.0012)	-0.0030** (0.0013)	-0.0026** (0.0013)	-0.0032*** (0.0013)	-0.0029** (0.0012)
N (math scores)	103,307,471	103,307,274	103,307,344	103,298,159	103,298,159

Notes: This table examines robustness for our event study estimates of the direct and indirect effects of exposure to household incarceration. Panel A presents estimates of δ from the direct event study regression 4, and Panel B presents estimates of δ from the indirect event study regression 6. Column (C) presents estimates from our benchmark specification, which replicates the results in column (C) of Tables 6 and 7. Our benchmark specification includes offense class at conviction (o) \times academic cohort (g) \times disposition year (τ) \times calendar year (t) fixed effects, which we denote by $\lambda_{og\tau(i)t}$. The other columns of this table present results using different definitions of this fixed effect term. Column (A) excludes interactions for offenses, o , and thus the fixed effects are $\lambda_{g\tau(i)t}$. Column (B) defines offenses, o , based on the offense class at the time of arrest (rather than at conviction). Column (D) defines offenses, o , by the four-digit offense code at conviction rather than the offense class at conviction. Column (E) is identical to column (D) except we add a covariate for the defendant's prior points at conviction interacted with a linear term for years since disposition, $t - \tau(i)$.

Parentheses contain standard errors clustered at the defendant (Panel A) and school (Panel B) levels.

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

TABLE A7: Judge stringency robustness tests

Dependent variable	(A) Benchmark	(B) Judge × year stringency	(C) Strin. from court × yr × offense residuals	(D) 4-digit offense code controls	(E) Felonies only
Panel A. Direct effects of a household incarceration					
Active sentence	0.0338*** (0.0033)	0.0334*** (0.0028)	0.0216*** (0.0016)	0.0280*** (0.0035)	0.0533*** (0.0090)
Math score	-0.0056 (0.0059)	-0.0086 (0.0055)	-0.0061* (0.0032)	-0.0020 (0.0075)	-0.0213* (0.0121)
English score	-0.0120** (0.0059)	-0.0122** (0.0053)	-0.0086*** (0.0032)	-0.0115 (0.0077)	-0.0229* (0.0125)
Days absent	0.0792* (0.0454)	-0.0265 (0.0454)	0.0023 (0.0279)	0.0321 (0.0551)	0.3058** (0.1196)
Any suspension	0.0033** (0.0015)	0.0028* (0.0014)	0.0019** (0.0008)	0.0037** (0.0017)	0.0146*** (0.0034)
Number of suspension days	0.0215 (0.0221)	0.0121 (0.0210)	0.0126 (0.0127)	0.0201 (0.0264)	0.0595 (0.0542)
Fighting incident	0.0014* (0.0008)	0.0006 (0.0007)	0.0007* (0.0004)	0.0011 (0.0009)	0.0014 (0.0018)
N (math scores)	295,905	285,192	303,712	294,331	84,431
Panel B. Indirect effects of a household incarceration on classmates					
Active sentence	0.0310*** (0.0023)	0.0298*** (0.0023)	0.0196*** (0.0015)	0.0258*** (0.0024)	0.0471*** (0.0076)
Math score	-0.0005 (0.0019)	-0.0021 (0.0021)	-0.0019* (0.0011)	-0.0024 (0.0022)	-0.0057 (0.0039)
English score	-0.0014 (0.0016)	-0.0029* (0.0017)	-0.0017* (0.0009)	-0.0036* (0.0019)	-0.0047 (0.0032)
N (math scores)	32,120,474	30,910,073	32,991,947	32,120,239	9,048,286

Notes: This table examines robustness for our judge stringency estimates of the direct and indirect effects of exposure to a household incarceration. Panel A presents reduced-form estimates of γ from the direct judge stringency regression 5. Panel B presents reduced-form estimates of γ from the indirect judge stringency regression 8. Column (A) presents estimates from our benchmark specification, which replicates the results in column (D) of Tables 6 and 7. For column (B), we compute stringency at the judge × year level rather than at the judge level (as in our benchmark). Column (C) is identical to column (B) except we compute stringency using a different method of residualizing. For column (C), we regress an indicator for an active sentence on dummies for court × year × offense class (rather than dummies for the structured sentencing grid) in our leave-out sample and then average the residuals from this regression at the judge × year level. Column (D) is identical to column (A) except that in the fixed effect term, θ_{ocr} , we define offenses, o , using the four-digit offense code rather than the offense class. Column (E) estimates our benchmark specification using only defendants who were charged with a felony, for which there is more variation in active sentencing. Parentheses contain standard errors clustered at the judge (Panel A) and school (Panel B) levels.

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

TABLE A8: Direct impacts on neighborhood quality and criminal activity

Dependent variable	(A)	(B)	(C)	Judge stringency estimate	
	Mean	OLS coef. on active sentence	Event study estimate	Reduced form (SD units)	IV for active sentence
Neighborhood quality and student criminal activity					
Neighborhood SES percentile	0.454	-0.024*** (0.004)	0.003* (0.002)	0.003 (0.002)	0.086 (0.066)
Student has criminal charge	0.004	0.002 (0.002)	-0.002 (0.001)	-0.000 (0.001)	-0.012 (0.016)
Judge stringency sample	Y	Y		Y	Y
Event study sample			Y		
Years, t , relative to case disposition, $\tau(i)$	$t \geq \tau(i)$	$t \geq \tau(i)$	All t	$t \geq \tau(i)$	$t \geq \tau(i)$
N (math scores)	296,242	296,242	699,084	296,242	296,242
# students	118,416	118,416	128,829	118,416	118,416
<i>F</i> -statistics:					
First stage				103.0	
Defendant balance test				0.9	
Student balance test				1.1	

Notes: This table presents estimates of the direct impacts of a household incarceration using our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our direct judge stringency sample (column (B) in Appendix Table A5). Column (C) includes students in our direct event study sample (column (A) in Appendix Table A5). See Appendix C.5 for details on our samples. Each row corresponds to a separate regression using the dependent variable listed in the first column. Regressions are at the student \times year level.

Column (A) displays means of each outcome variable measured in calendar years, t , in or after the disposition year of the defendant’s case, $\tau(i)$. Column (B) shows OLS estimates of β from Equation 3 using outcomes measured in years $t \geq \tau(i)$, and column (C) displays estimates of δ from the event study specification 4 using all years t . Column (D) shows reduced-form judge stringency estimates of γ from Equation 5 for years $t \geq \tau(i)$, with estimates normalized to represent a one standard deviation increase in judge stringency. Column (E) shows estimates of β from the IV specification based on Equations 3 and 5 using years $t \geq \tau(i)$. Regressions in columns (B), (D), and (E) include court \times year \times offense class dummies and defendant and student characteristics (age at disposition, gender, race dummies, student socioeconomic status, and missing values of each covariate). Parentheses contain standard errors clustered at the defendant (column (C)) and judge (columns (B), (D), and (E)) levels. The bottom of the table reports *F*-statistics from the first stage and joint significance tests for defendant and student characteristics for the specification in column (D); see Appendix Table B1 for details on the balance tests.

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

TABLE A9: Decomposition of the gradient between test scores and school incarceration exposure

	(A)	(B)	(C)	(D)
Panel A. Students and incarceration exposure				
	Individual	School/grade	School	
Mean # students	1.0	192.2	715.2	
Mean # HH incarcerations/year	0.023	4.5	16.7	
Panel B. Estimates from paper				
	Effect of 1 HH incarceration in event study	Effect of 1 log point in HH incarcerations/ student in school		
	Direct	School/grade (indirect)	Raw gradient	With controls
Math score	-0.015	-0.004	-0.205	-0.104
English score	-0.012	-0.003	-0.208	-0.094
Panel C. Effect of 1 HH incarceration in school at mean				
	Direct	School/grade (indirect)	Raw gradient	With controls
Math score	-0.00002	-0.0011	-0.0120	-0.0061
% of raw gradient	0.2%	9%	100%	51%
English score	-0.00002	-0.0007	-0.0121	-0.0055
% of raw gradient	0.1%	6%	100%	45%

Notes: This table uses event study estimates from the paper to decompose the gradient between test scores and school incarceration exposure into direct, indirect, and residual channels. Panel A reports summary statistics at the individual (column (A)), school/grade (column (B)), and school (column (C)) levels. We report the mean number of students in each group across all North Carolina public school students in our data from 2010–2014. We also report the mean number of household incarcerations in each group. For example, the average public school student experienced 0.023 household incarcerations in a typical year but were in the same school/grade as 4.5 students with a household incarceration.

Panel B summarizes estimates from the paper. Column (A) displays event study estimates of the direct impacts of a household incarceration on math and English scores (column (C) of Table 6). Column (B) displays event study estimates of the indirect impacts of a household incarceration on math and English scores (column (C) of Table 7). Column (C) displays estimates of the raw gradient between school average test scores and log household incarcerations (column (A) of Table 2, Panel A). Column (D) displays estimates of this gradient with demographic, county, and year controls (column (C) of Table 2, Panel A).

Panel C combines the statistics from Panels A–B to estimate the effects of one additional household incarceration in a school at the mean. We provide examples of each calculation for math scores; the calculations for English scores are analogous. Column (A) estimates the direct impact of one household incarceration on school mean math scores; this equals the direct event study estimate (–0.015) times the proportion of students in the school who are affected by this incarceration (1/715.2). Column (B) estimates the indirect impacts of one household incarceration on school mean math scores through the channel of school/grade peers; this equals the indirect event study estimate (–0.004) times the proportion of students in the school who are in the same grade as the directly impacted student (192.2/715.2). Column (C) reports the effect of one household incarceration on the raw gradient at the mean; this equals the raw gradient (–0.205) times the log point change of 1 incarceration at the mean ($\log(16.7 + 1) - \log(16.7)$). Column (D) reports the effect of one household incarceration on the gradient with controls at the mean; this equals the gradient with controls (–0.104) times the log point change of 1 incarceration at the mean ($\log(16.7 + 1) - \log(16.7)$). Panel C also reports each component’s percentage of the raw gradient, which is the estimate in each column divided by the estimate in column (C).

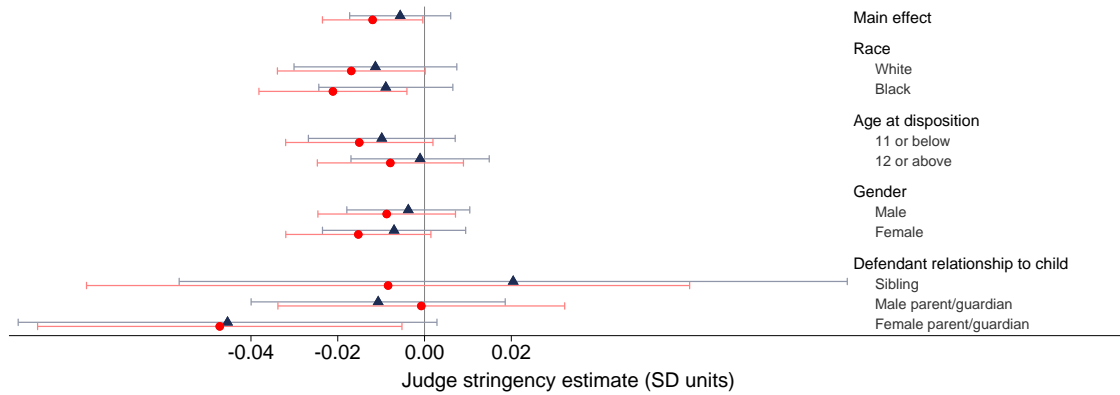
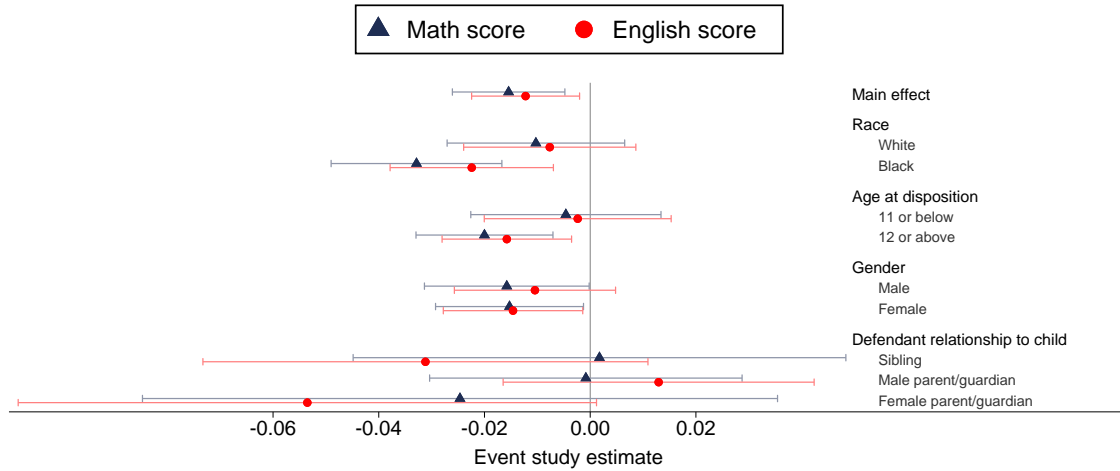
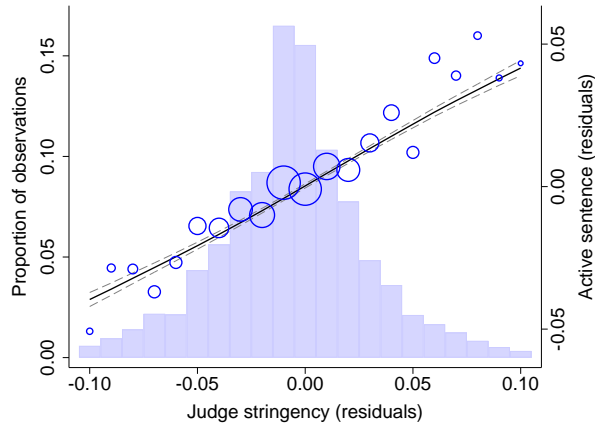
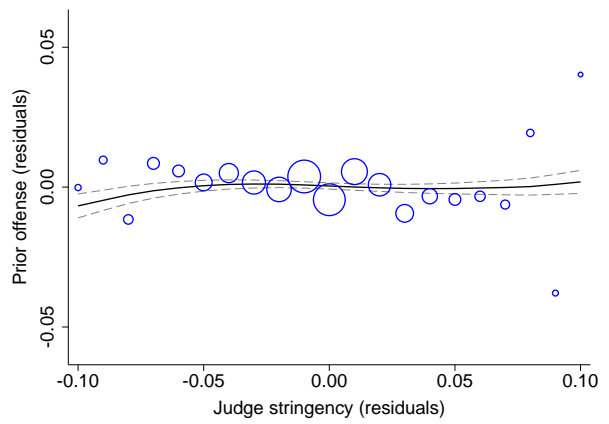


FIGURE A1: Heterogeneity in the direct effects of household incarceration

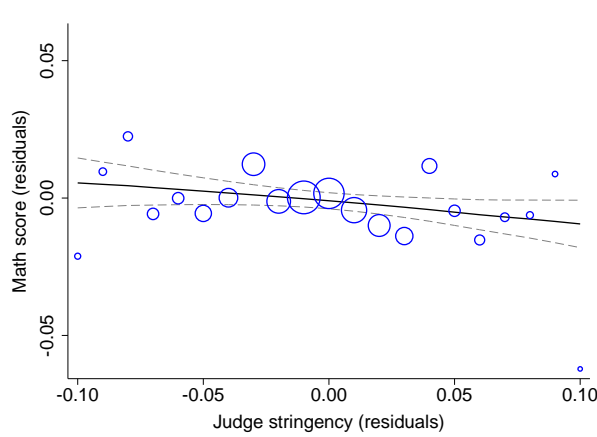
Notes: This figure displays heterogeneity in the direct impacts of a household incarceration on student test scores. Panel A presents event study estimates using our direct exposure sample from column (A) in Appendix Table A5. Panel B presents reduced-form judge stringency estimates using our direct exposure sample from column (B) in Appendix Table A5, normalized to represent a one standard deviation increase in judge stringency. The main effects replicate the estimates from columns (C)–(D) in Panel B of Table 6. All other coefficients come from estimating the same regressions in the subsamples listed on the right side of each panel. We define parent/guardians as defendants who are 20–40 years older than the student, and we define siblings as defendants who are 1–10 years older than the student. The outcome variables are math (blue triangles) and English (red circles) scores in standard deviation units, defined as in Table 1. Lines depict 95 percent confidence intervals using standard errors clustered at the defendant (Panel A) and judge (Panel B) levels.



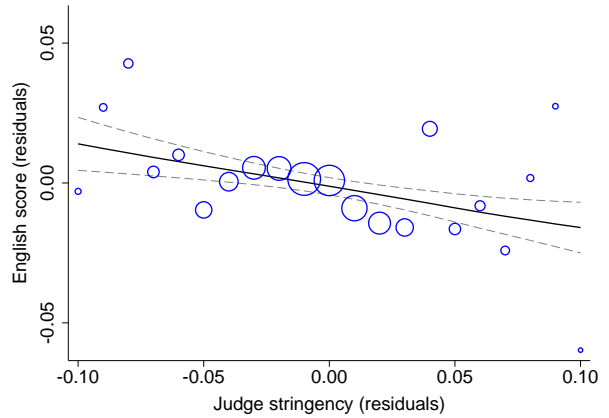
Panel A. Active sentence



Panel B. Prior offense



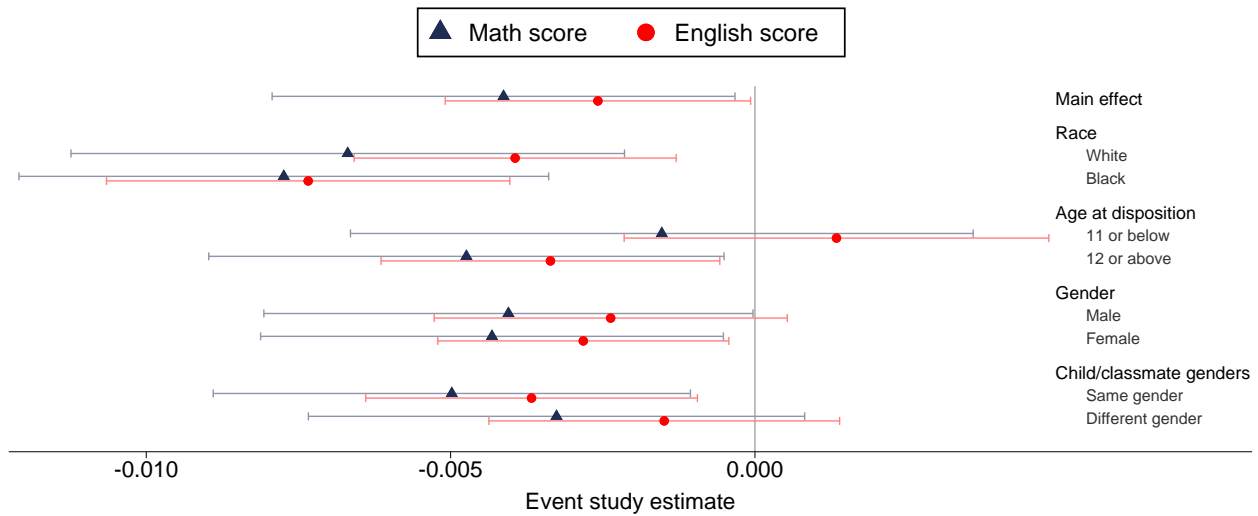
Panel C. Math score (directly impacted child)



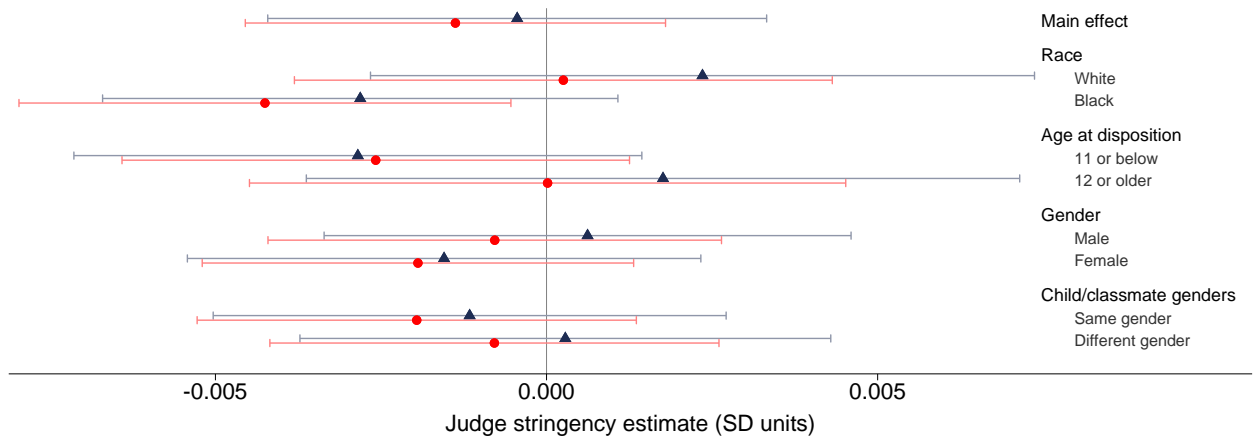
Panel D. English score (directly impacted child)

FIGURE A2: Reduced-form judge stringency effects

Notes: This figure shows the distribution of judge stringency and its relationship with defendant and student outcomes. Each panel depicts the relationship between outcome residuals (y -axis) and judge stringency residuals (x -axis). We compute these residuals from a regression of each variable on court \times year \times offense class dummies. The sample includes students in our direct judge stringency sample (column (B) of Appendix Table A5) and their linked defendants. The outcome variables are: an indicator equal to one if the defendant received an active sentence (Panel A); an indicator equal to one if the defendant had a prior offense in our data (Panel B); the directly impacted student's average math score in all years in or after the disposition (Panel C); and the directly impacted student's average English score in all years in or after the disposition (Panel D). Circles show means of each variable in 0.01 bins of judge stringency residuals, with sizes proportional to the number of observations. The solid line depicts predicted values from a local linear regression of outcome residuals on judge stringency residuals. Dashed lines plot 95 percent confidence intervals. In Panel A, bars show the distribution of judge stringency residuals (weighted by the number of student observations) in 0.01 unit bins (left y -axis). The graphs exclude judge stringency residuals below -0.1 and above 0.1 (2 percent of the sample).



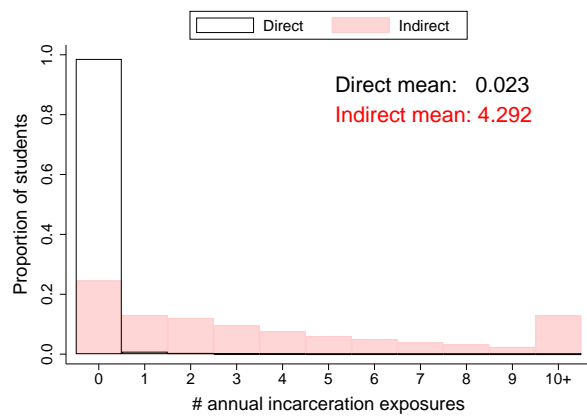
Panel A. Event study



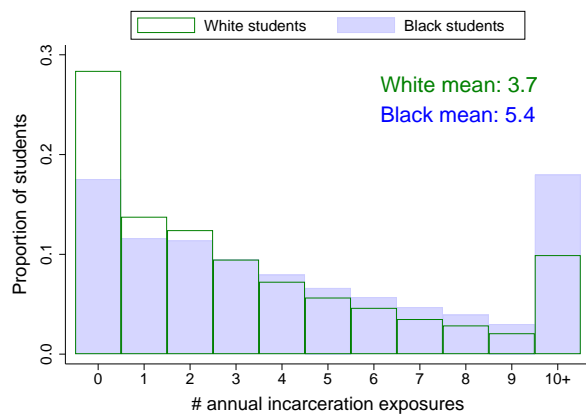
Panel B. Judge stringency

FIGURE A3: Heterogeneity in the indirect effects of household incarceration

Notes: This figure displays heterogeneity in the indirect impacts of a household incarceration on classmates' test scores. Panel A presents event study estimates using our indirect exposure sample from column (D) in Appendix Table A5. Panel B presents reduced-form judge stringency estimates using our direct exposure sample from column (E) in Appendix Table A5, normalized to represent a one standard deviation increase in judge stringency. The main effects replicate the estimates from columns (C)–(D) in Panel B of Table 7. All other coefficients come from estimating the same regressions in the subsamples listed on the right side of each panel. Characteristics for this heterogeneity are those of the indirectly impacted student except for peer/child gender, which is based on both the directly and indirectly impacted students. The outcome variables are math (blue triangles) and English (red circles) scores in standard deviation units, defined as in Table 1. In both panels, lines depict 95 percent confidence intervals using standard errors clustered at the school level.



Panel A. Direct and indirect exposure



Panel B. Indirect exposure by race

FIGURE A4: Distributions of direct and indirect exposure to household incarceration

Notes: This figure displays histograms of direct and indirect exposure to household incarceration. Direct exposure means that a student has a household member who receives an active sentence in a given year. Indirect exposure means that a student is in the same school and grade as a child with a direct exposure in the year of the defendant's disposition. The sample includes all North Carolina public school students in 2010–2014 (column (A) of Table 1). Panel A shows the distributions of annual direct and indirect exposures, and Panel B shows the distributions of indirect exposures for white and Black students. In both panels we group 10 or more exposures into a single category.

B The Role of Judges in the North Carolina Court System

B.1 Court Structure and Judge Turnover

North Carolina's criminal court system is organized into two divisions: the Appellate Division, and the Trial Division comprising the superior and division courts.³⁰ We focus on district and superior courts, which hear new criminal cases. Superior courts have jurisdiction over all felony cases that are heard by trial, as well as misdemeanor cases that are appealed from district court. District courts have jurisdiction over all new misdemeanor cases. Felony cases often begin in District Court for pre-trial proceedings; class H or class I felonies may also be resolved in district court with a guilty plea. Superior court trials are presided over by a judge with a jury of twelve, whereas district court trials are always held without jury.

Criminal cases are heard in the county court where they are filed, but both district and superior Courts are aggregated for electoral and administrative purposes. District courts are grouped into 41 different "districts," as shown in Panel A of Figure B3. These districts contain either one large county or several small counties. District court judges are elected to these districts and hold court in all counties that comprise their district. Superior courts are grouped into 48 districts and 5 divisions, as shown in Panel A of Figure B4. Superior court judges are also elected at the district level, but the state constitution requires that superior court judges rotate to different districts within their division every six months.

Turnover in the judges who work in each county court results from elections, retirements, promotions, and the superior court rotation system. District and superior court judges are elected in biannual November elections. District court judges are elected for four-year terms, and Superior Court judges serve eight-year terms. During our sample period (2010–2014), all judicial elections were non-partisan. If judges leave during the middle of their term, the governor appoints a replacement until the next election (superior court judges) or until the end of the departing judge's term (district court judges), at which point the seat is put up for election. Our analysis also exploits variation in judges' caseloads conditional on serving in a county court, which arises from scheduling constraints, variation in trial length, vacations, and other factors.

³⁰ For details on the North Carolina court structure, see: <https://www.nccourts.gov/courts>.

B.2 Timeline of Court Process

Individuals who are arrested and charged in the state of North Carolina are typically brought in front of a magistrate judge within 48 hours of the arrest. During the arrangement, the magistrate makes a preliminary determination whether individuals may be released on bond (and for what amount); whether they will be released without bail (personal recognizance); or whether pretrial release is not granted. The magistrate also determines the date of the preliminary hearing in front of the district court judge, which typically happens within 1–3 days.

The preliminary hearing is intended to explain the charges to the defendant and arrange for legal council. If required, the district court judge will appoint a defense attorney—who may be either a public defender or an appointed local private attorney if no public defender is available. Misdemeanor cases are typically resolved at the district court level with the final sentence determined, without a jury, by the district court judge.

Since the 1994 Structured Sentencing Act, judicial sentencing has been limited under a grid system designed to limit the extent of judicial discretion, as in Figure B2. These establish minimum and maximum ranges for sentencing based on offense severity and the number of prior convictions. They also establish the different sentencing options available to judges: 1) community punishment (probation), 2) intermediate punishment (probation with additional conditions³¹), and 3) active punishment—incarceration in prison or jail. An active sentence necessarily implies judicial incarceration for at least the minimum sentence length.³² Sentencing may also take into consideration time served in jail during the pretrial stage.

Felony cases feature a more complicated process involving the superior court. Defendants in these cases follow preliminary arrangements and hearings at the district court level. However, the District Court judge will then schedule a hearing requiring the District Attorney to produce probable cause within 5–15 days. If probable cause is found, the case is moved to Superior Court and a grand jury hearing is held to determine whether there is sufficient evidence to indict the defendant.

³¹ This can include: 1) probation combined with incarceration spells (either beginning with a period of incarceration, or periodic spells of confinement) 2) additional monitoring and treatment for substance abuse, 3) community service, or 4) special electronic monitoring through electronic or satellite methods.

³² From the 2009 Structured Sentencing training manual (Spainhour and Katzenelson (2009), p. 28): “Active Punishment G.S. 15A-1340.11(1) An active punishment requires that the offender be sentenced to the custody of the Department of Correction to serve the minimum and up to the maximum sentence imposed by the court.”

The appointed Superior Court Judge sets bond amounts and trial dates in this case, manages the case, and takes the defendant plea. The trial will typically proceed with a jury, and the Superior Judge will determine the sentencing following the structured sentencing grid in Figure B2. Felony sentence levels also vary within three sets of ranges: mitigated, presumptive (the majority of cases), and aggravated, reflecting the severity of the offense. Felony prior records operate on a point system in which misdemeanors receive one point, and prior felonies receive 2–10 points depending on the precise offense.

B.3 Assignment of Cases to Judges

The assignment of cases to judges is determined by county clerks. At the district court level, the senior judge, in conjunction with the county clerk, determines an assignment of judges to specific courtrooms specializing in different categories of cases.³³ A similar rotational process happens at the superior court level. The result is that judges rotate across hearing cases of different types, and oversee cases within that specific domain during the periods in which they are active.³⁴

For our judge randomization strategy to be effective, we require that judges vary in sentence severity and that criminal defendants (and so the children living with these defendants) face quasi-random variation in exposure to different judges. Figure B6 shows the persistence of judge stringency over time, consistent with the idea that sentencing severity is a relatively fixed judicial trait. In Table B1 we show that exposure to judges with higher sentencing severity is not associated with other background characteristics of either defendants or students (columns (B) and (D)). This is in contrast to the strong relationship between active sentences and these characteristics (columns (A) and (C)).

³³ For example, the rotation in Buncombe County is weekly and can be found for period of December 28, 2020–July 2, 2021 at <https://www.nccourts.gov/assets/inline-files/Buncombe-DCJ-Rotation-Dec-28-2020-July-2-2021.pdf>.

³⁴ The North Carolina superior court judges' Benchbook discusses the nature of case sessions in more detail: https://benchbook.sog.unc.edu/sites/default/files/pdf/Out%20f%20Term_Out%20f%20Session_Out%20of%20County.pdf.

*****Effective for Offenses Committed on or after 12/1/13*****

MISDEMEANOR PUNISHMENT CHART

CLASS	PRIOR CONVICTION LEVEL			
	I	II		III
	No Prior Convictions	One to Four Prior Convictions		Five or More Prior Convictions
A1	C/I/A 1 - 60 days	C/I/A 1 - 75 days		C/I/A 1 - 150 days
1	C 1 - 45 days	C/I/A 1 - 45 days		C/I/A 1 - 120 days
2	C 1 - 30 days	C/I 1 - 45 days		C/I/A 1 - 60 days
3	C Fine Only* 1 - 10 days	One to Three Prior Convictions	Four Prior Convictions	C/I/A 1 - 20 days
		C Fine Only* 1 - 15 days	C/I 1 - 15 days	

***Unless otherwise provided for a specific offense, the judgment for a person convicted of a Class 3 misdemeanor who has no more than three prior convictions shall consist only of a fine.**

A – Active Punishment I – Intermediate Punishment C – Community Punishment
Cells with slash allow either disposition at the discretion of the judge

FIGURE B1: Range of outcomes under North Carolina’s structured sentencing: Misdemeanors

Notes: This matrix illustrates the range of judicial discretion for misdemeanor cases in North Carolina under structured sentencing laws during the sample period. This document is taken from the “Citizen’s Guide to Structured Sentencing,” available from the North Carolina Courts at <https://www.nccourts.gov/documents/publications/citizens-guide-to-structured-sentencing>. Each cell corresponds to a combination of offense class and prior record level. Within each cell, roman numerals indicate whether sentencing outcomes of “A” (active sentencing), “I” (intermediate punishment), and “C” (community punishment) are available to judges. Ranges of numbers indicate the days of minimum sentencing available to judges.

***** Effective for Offenses Committed on or after 10/1/13 *****

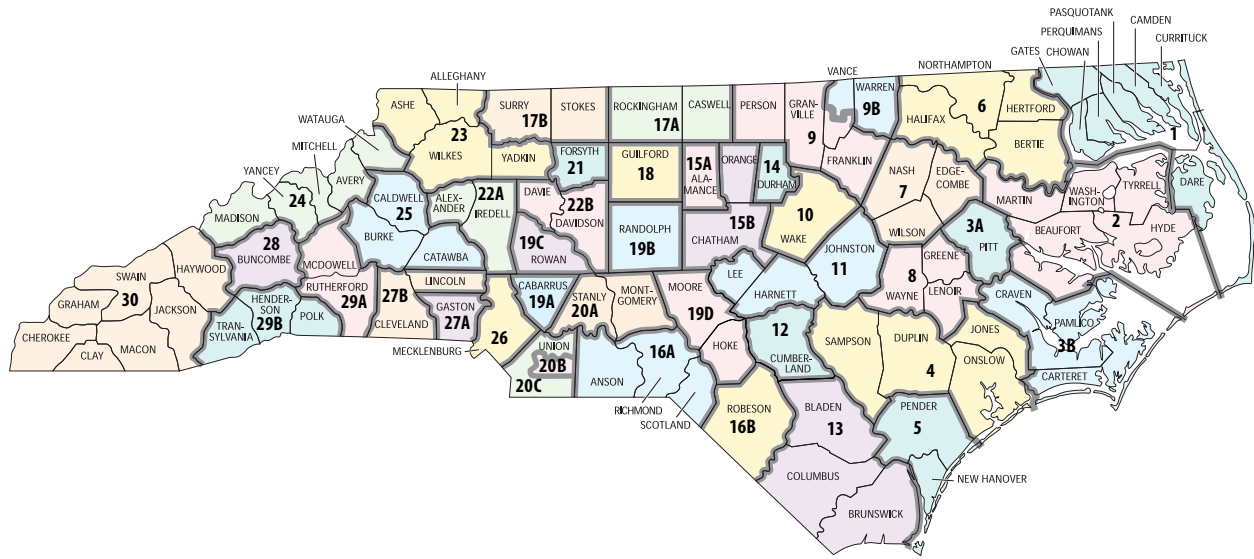
**FELONY PUNISHMENT CHART
PRIOR RECORD LEVEL**

		I 0-1 Pt	II 2-5 Pts	III 6-9 Pts	IV 10-13 Pts	V 14-17 Pts	VI 18+ Pts	
		Death or Life Without Parole						
		Defendant Under 18 at Time of Offense: Life With or Without Parole						
OFFENSE CLASS	A							DISPOSITION
	B1	A	A	A	A	A	A	<i>Aggravated Range</i>
		<i>240 - 300</i>	<i>276 - 345</i>	<i>317 - 397</i>	<i>365 - 456</i>	<i>Life Without Parole</i>	<i>Life Without Parole</i>	
		192 - 240	221 - 276	254 - 317	292 - 365	336 - 420	386 - 483	PRESUMPTIVE RANGE
		<i>144 - 192</i>	<i>166 - 221</i>	<i>190 - 254</i>	<i>219 - 292</i>	<i>252 - 336</i>	<i>290 - 386</i>	<i>Mitigated Range</i>
	B2	A	A	A	A	A	A	
		<i>157 - 196</i>	<i>180 - 225</i>	<i>207 - 258</i>	<i>238 - 297</i>	<i>273 - 342</i>	<i>314 - 393</i>	
		125 - 157	144 - 180	165 - 207	190 - 238	219 - 273	251 - 314	
		<i>94 - 125</i>	<i>108 - 144</i>	<i>124 - 165</i>	<i>143 - 190</i>	<i>164 - 219</i>	<i>189 - 251</i>	
	C	A	A	A	A	A	A	
<i>73 - 92</i>		<i>83 - 104</i>	<i>96 - 120</i>	<i>110 - 138</i>	<i>127 - 159</i>	<i>146 - 182</i>		
58 - 73		67 - 83	77 - 96	88 - 110	101 - 127	117 - 146		
	<i>44 - 58</i>	<i>50 - 67</i>	<i>58 - 77</i>	<i>66 - 88</i>	<i>76 - 101</i>	<i>87 - 117</i>		
D	A	A	A	A	A	A		
	<i>64 - 80</i>	<i>73 - 92</i>	<i>84 - 105</i>	<i>97 - 121</i>	<i>111 - 139</i>	<i>128 - 160</i>		
	51 - 64	59 - 73	67 - 84	78 - 97	89 - 111	103 - 128		
	<i>38 - 51</i>	<i>44 - 59</i>	<i>51 - 67</i>	<i>58 - 78</i>	<i>67 - 89</i>	<i>77 - 103</i>		
E	I/A	I/A	A	A	A	A		
	<i>25 - 31</i>	<i>29 - 36</i>	<i>33 - 41</i>	<i>38 - 48</i>	<i>44 - 55</i>	<i>50 - 63</i>		
	20 - 25	23 - 29	26 - 33	30 - 38	35 - 44	40 - 50		
	<i>15 - 20</i>	<i>17 - 23</i>	<i>20 - 26</i>	<i>23 - 30</i>	<i>26 - 35</i>	<i>30 - 40</i>		
F	I/A	I/A	I/A	A	A	A		
	<i>16 - 20</i>	<i>19 - 23</i>	<i>21 - 27</i>	<i>25 - 31</i>	<i>28 - 36</i>	<i>33 - 41</i>		
	13 - 16	15 - 19	17 - 21	20 - 25	23 - 28	26 - 33		
	<i>10 - 13</i>	<i>11 - 15</i>	<i>13 - 17</i>	<i>15 - 20</i>	<i>17 - 23</i>	<i>20 - 26</i>		
G	I/A	I/A	I/A	I/A	A	A		
	<i>13 - 16</i>	<i>14 - 18</i>	<i>17 - 21</i>	<i>19 - 24</i>	<i>22 - 27</i>	<i>25 - 31</i>		
	10 - 13	12 - 14	13 - 17	15 - 19	17 - 22	20 - 25		
	<i>8 - 10</i>	<i>9 - 12</i>	<i>10 - 13</i>	<i>11 - 15</i>	<i>13 - 17</i>	<i>15 - 20</i>		
H	C/I/A	I/A	I/A	I/A	I/A	A		
	<i>6 - 8</i>	<i>8 - 10</i>	<i>10 - 12</i>	<i>11 - 14</i>	<i>15 - 19</i>	<i>20 - 25</i>		
	5 - 6	6 - 8	8 - 10	9 - 11	12 - 15	16 - 20		
	<i>4 - 5</i>	<i>4 - 6</i>	<i>6 - 8</i>	<i>7 - 9</i>	<i>9 - 12</i>	<i>12 - 16</i>		
I	C	C/I	I	I/A	I/A	I/A		
	<i>6 - 8</i>	<i>6 - 8</i>	<i>6 - 8</i>	<i>8 - 10</i>	<i>9 - 11</i>	<i>10 - 12</i>		
	4 - 6	4 - 6	5 - 6	6 - 8	7 - 9	8 - 10		
	<i>3 - 4</i>	<i>3 - 4</i>	<i>4 - 5</i>	<i>4 - 6</i>	<i>5 - 7</i>	<i>6 - 8</i>		

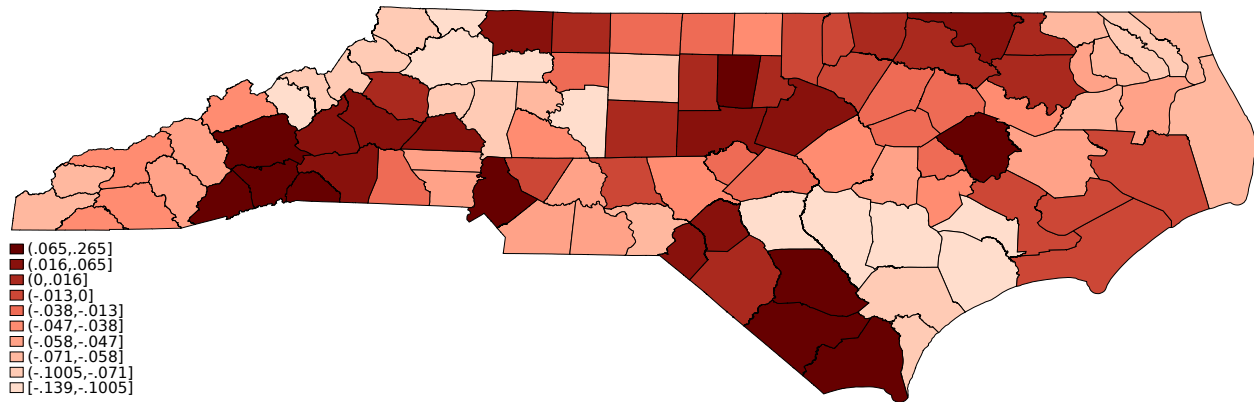
A – Active Punishment I – Intermediate Punishment C – Community Punishment
 Numbers shown are in months and represent the range of minimum sentences

FIGURE B2: Range of outcomes under North Carolina’s structured sentencing: Felonies

Notes: This matrix illustrates the range of judicial discretion for felony cases in North Carolina under structured sentencing laws during the sample period. This document is taken from the “Citizen’s Guide to Structured Sentencing,” available from the North Carolina Courts at <https://www.nccourts.gov/documents/publications/citizens-guide-to-structured-sentencing>. Each cell corresponds to a combination of offense class and prior record level. Within each cell, roman numerals indicate whether sentencing outcomes of “A” (active sentencing), “I” (intermediate punishment), and “C” (community punishment) are available to judges. Ranges of numbers indicate the months of minimum sentencing available to judges for cases in the presumptive, aggravated, and mitigated ranges.



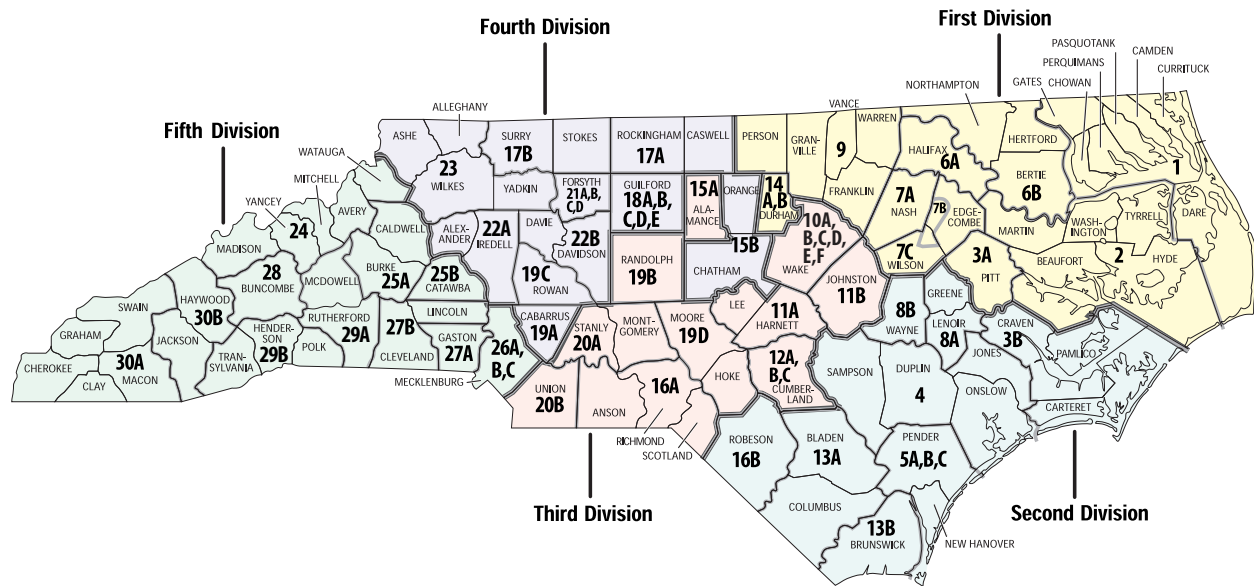
Panel A. District court districts (Effective January 1, 2019)



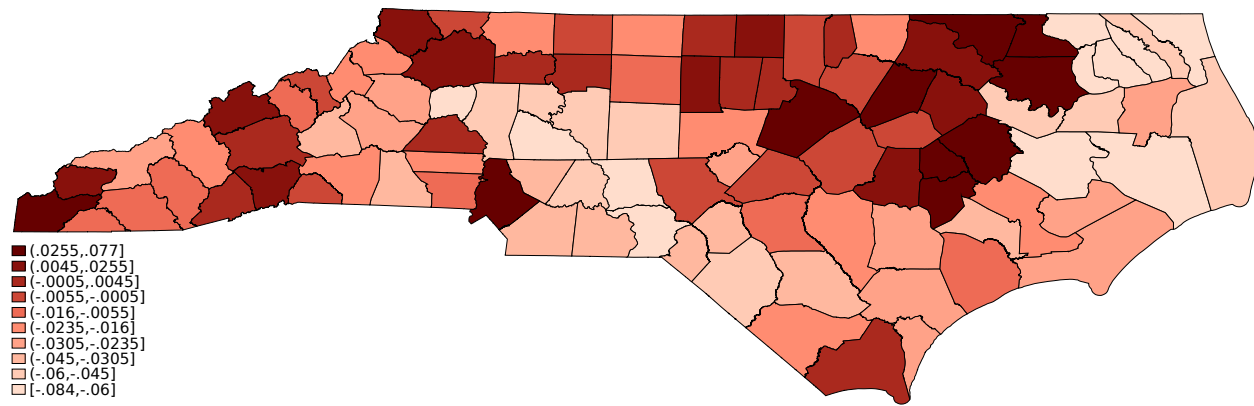
Panel B. Mean stringency for district court judges (2010)

FIGURE B3: North Carolina district courts

Notes: Panel A shows judicial district court boundaries, obtained from <https://www.sog.unc.edu/resource-series/judicial-maps>. Panel B shows mean stringency for district court judges in 2010, which we compute following the methodology described in Section 3.1.



Panel A. Superior court districts (Effective January 1, 2019)



Panel B. Mean stringency for superior court judges (2010)

FIGURE B4: North Carolina superior courts

Notes: Panel A shows judicial superior court boundaries, obtained from <https://www.sog.unc.edu/resource-series/judicial-maps>. Panel B shows mean stringency for superior court judges in 2010, which we compute following the methodology described in Section 3.1.

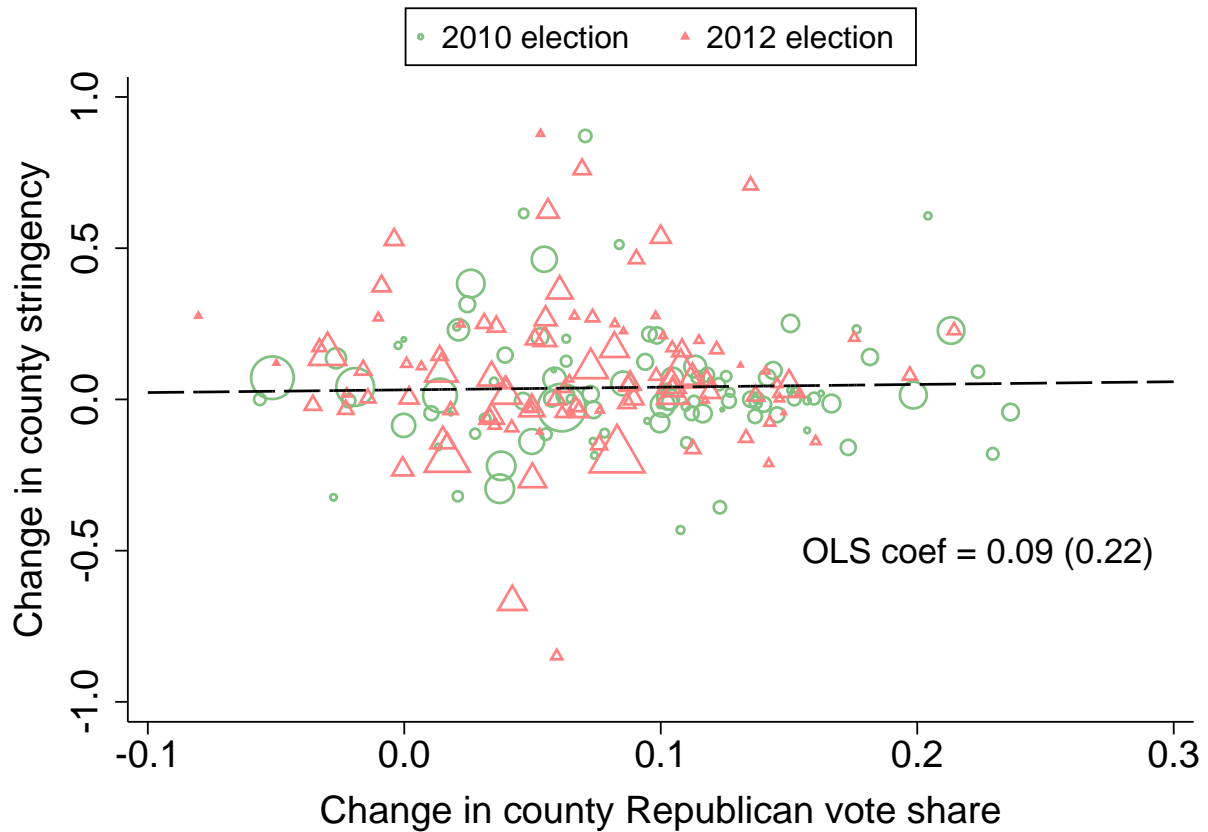
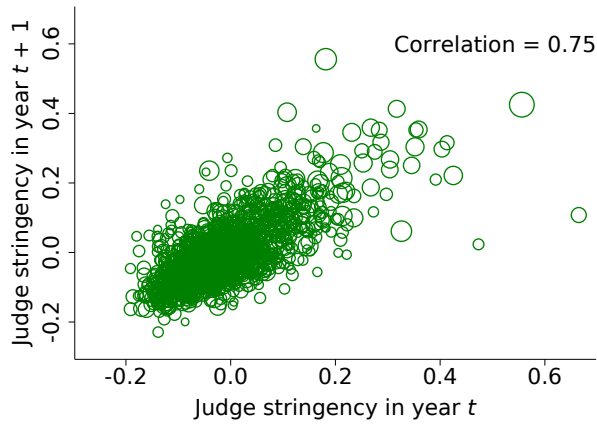
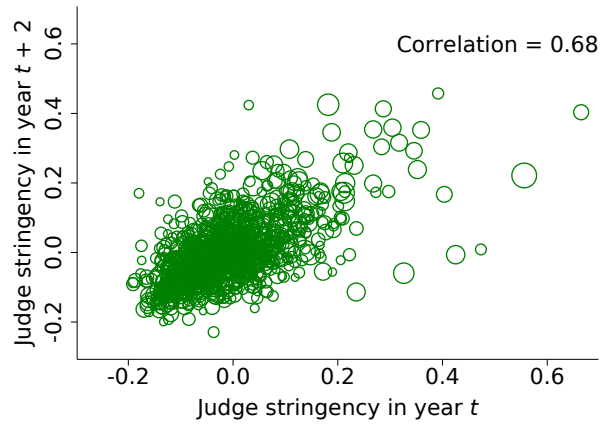


FIGURE B5: Changes in county stringency and county-level Republican vote share

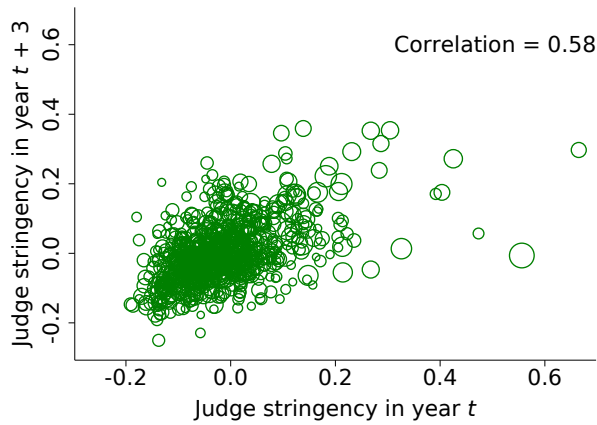
Notes: This figure plots the relationship between changes in county stringency and the county-level Republican vote share. The x -axis shows the change in the Republican vote share between years t and $t - 4$, where t includes the November 2010 election (green circles) and the November 2012 election (red triangles). We use a four-year deviation because North Carolina district court judges serve four-year terms, so the x -axis represents the change in voting behavior from the last election cycle for these judges. To compute the Republican vote share, we use all state house, state senate, and U.S. congressional races that featured both a Republican and a Democratic candidate, and we exclude all third-party candidates. The election data come from the North Carolina State Board of Elections (<https://www.ncsbe.gov/results-data/election-results>). The y -axis plots the change in county stringency between years t and $t - 1$, where t includes 2011 and 2013. For this figure, we use calendar years (as opposed to academic years) to compute county stringency in order to align with the timing of judge turnover from elections. The dashed line shows the OLS relationship between the y - and x -axes. The OLS coefficient is 0.09 with a robust standard error of 0.22.



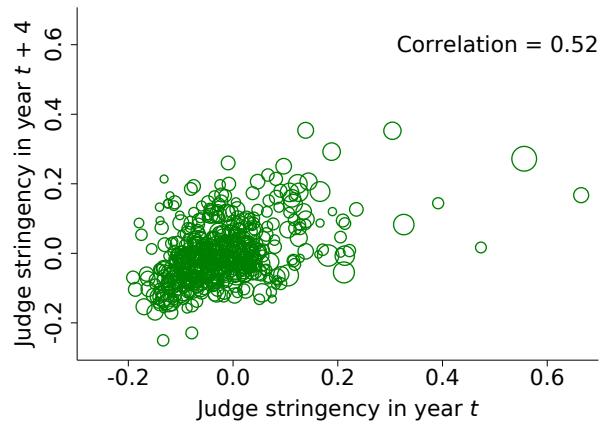
Panel A. Years t and $t + 1$



Panel B. Years t and $t + 2$



Panel C. Years t and $t + 3$



Panel D. Years t and $t + 4$

FIGURE B6: Persistence of individual judge stringency

Notes: This figure displays the persistence of judge stringency over time. For this figure, we compute judge stringency using the same methodology as described in Section 3.1, but we average active sentence residuals at the judge \times disposition year level. This gives a measure of stringency for each judge in each year in our sample (2010–2014). Panel A shows a scatterplot of each judge’s stringency in year t (x -axis) against the same judge’s stringency in the subsequent year, $t + 1$ (y -axis). The other panels are similar except the y -axis displays the judge’s stringency in years $t + 2$ (Panel B), $t + 3$ (Panel C), and $t + 4$ (Panel D). Each panel displays the correlation coefficient between the stringency measures in the two years.

TABLE B1: Judge stringency balance tests

Covariate	(A)	(B)	(C)	(D)
	Dependent variable in direct sample		Dependent variable in indirect sample	
	Active sentence	Judge stringency (SD units)	Active sentence	Judge stringency (SD units)
Panel A. Defendant/offense characteristics				
Male	0.040*** (0.003)	0.006 (0.005)	0.038*** (0.002)	0.008 (0.006)
Age at disposition	0.001*** (0.000)	-0.000 (0.000)	0.001*** (0.000)	-0.000 (0.000)
White	-0.031*** (0.005)	-0.002 (0.009)	-0.028*** (0.004)	-0.002 (0.008)
Black	-0.005 (0.005)	0.000 (0.011)	-0.002 (0.004)	0.002 (0.009)
Multiple offenses	0.026*** (0.003)	0.009 (0.006)	0.026*** (0.003)	0.015** (0.006)
Prior offense	0.014*** (0.003)	-0.002 (0.005)	0.015*** (0.003)	-0.002 (0.006)
Prior active sentence	0.231*** (0.010)	0.019** (0.009)	0.235*** (0.010)	0.017 (0.012)
<i>F</i> -statistic: All coefficients zero	97.8	1.1	136.9	1.3
Panel B. Student characteristics				
Male	0.001 (0.002)	-0.000 (0.003)	0.000 (0.000)	-0.000 (0.000)
Age at disposition	0.000 (0.000)	-0.001 (0.000)	-0.000 (0.000)	-0.000 (0.001)
White	-0.031*** (0.004)	-0.004 (0.006)	-0.006*** (0.001)	-0.001 (0.002)
Black	-0.003 (0.003)	-0.003 (0.007)	0.002** (0.001)	0.002 (0.002)
Economically disadvantaged	0.019*** (0.002)	0.006 (0.004)	0.003*** (0.001)	0.002* (0.001)
<i>F</i> -statistic: All coefficients zero	37.7	1.1	13.7	1.2
<i>N</i>	118,416	118,416	17,435,938	17,435,938
# students	118,416	118,416	1,454,577	1,454,577

Notes: This table presents balance tests for our judge stringency empirical strategy. Panel A tests for balance with respect to defendant/offense characteristics, and Panel B tests for balance with respect to student characteristics. Columns (A) and (C) display results from regressions of an active sentence indicator on the covariates listed in the first column; these columns show the OLS relationship between individual characteristics and active sentencing. Columns (B) and (D) instead use judge stringency as the dependent variable; these columns present our balance tests. Regressions in columns (A)–(B) are at the student level using students in our direct judge stringency sample (column (B) of Appendix Table A5). Regressions in columns (C)–(D) are at the directly impacted student \times classmate level using all classmates in our indirect judge stringency sample (column (E) of Appendix Table A5). All regressions include court \times year \times offense class dummies. The bottom of each panel displays *F*-statistics from joint significance tests for the coefficients on all covariates. Parentheses contain standard errors clustered at the judge (columns (A)–(B)) and school (columns (C)–(D)) levels.

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

C Empirical Appendix

This appendix provides details on our variable definitions, data sources, data cleaning, merge process, and analysis samples.

C.1 Definition of key variables and terms

- **Active sentence.** A binary indicator for the defendant receiving an Active Punishment from their criminal case. Throughout the paper, we use this term synonymously with “incarceration” because active sentences require that the defendant serves time in the custody of the Department of Corrections under North Carolina’s structured sentencing regulations. We define all other defendant outcomes as non-active, including Intermediate Punishments, Community Punishments, and verdicts of not guilty.
- **Any suspension.** A binary indicator for a student receiving any long-term suspension, short-term suspension, or expulsion during the academic year.
- **Calendar/school year.** For all outcomes in our paper, we define years based on the academic calendar (July through June) and refer to them by the year of the spring semester. For example, the year 2010 runs from July 1, 2009 through June 30, 2010. This aligns with the timing at which variables are measured in our education data. Thus we use the terms “calendar year” and “school year” interchangeably throughout the paper.
- **Cohort.** We define a student’s cohort as their expected year of high school graduation assuming on-time progression. This is a fixed student attribute that is determined by their grade level (K–12) in the first year they appear in the education data. For example, a student who first enrolls in kindergarten in 2000 is in the 2012 cohort.
- **County stringency.** County stringency is the county \times year level average of judge stringency (defined below). We compute a weighted average of judge stringencies using two different sets of weights. Our “actual caseload” measure uses weights that equal the total number of criminal cases that the judge heard in a given county \times year. Our “fixed caseload” measure includes only judges who served in a given county \times year, and we use as weights the judge’s

average number of criminal cases that they heard in that county across all years. For this we define serving in a given county \times year as hearing 20 or more criminal cases.

- **Court.** We define a court as a judge type \times county. Judge type is either superior or district court as determined by our classification of judges in the criminal offense records (see Appendix C.3.2). County is the county where the criminal case was filed in the criminal case records.
- **Days absent.** The student's total number of days of absence during the academic year.
- **Disposition year.** The year in which the defendant's case was decided in the criminal offense records. We define disposition years based on the academic calendar (July through June) as discussed above.
- **English score.** English scores include scores on both end-of-grade 3–8 reading exams and the end-of-course high school English I exam, which is most commonly taken in 9th or 10th grade. We standardize these scores to be mean zero and standard deviation one in the full population of test takers in each year. If the student took an exam multiple times in the same academic year, we use their first score. If the student took an end-of-grade and the English I exam in the same year, we use the average of their two standardized scores.
- **Exposure to incarceration (direct and indirect).** We define a student as *directly* exposed to an incarceration if they are linked to a criminal defendant who receives an active sentence; see Appendix C.4 for details on the linking of students and defendants. We define a student as *indirectly* exposed if they attended the same school and grade as a directly-exposed student in the year of the defendant's disposition. Throughout the paper we refer to indirectly-exposed children as "classmates" of the directly-exposed child since these students typically move through the school system together, and thus often share classes.
- **Fighting incident.** A binary indicator for a student having any fighting incident during the academic year, defined by the student offense code 024.
- **Judge stringency.** We define a measure of judicial stringency that captures a judge's average tendency to impose an active sentence among the convictions they oversee (versus an

intermediate or community punishment). We use this measure in our judge turnover analysis in Section 3, and also in our analysis of mechanisms in Sections 4–5. Our measure of judge stringency is similar throughout the paper, but the implementation details vary to be appropriate for the two analyses.

For Section 3, we begin by running a regression of an active sentence indicator on dummies for the structured sentencing grid. The sample for this regression includes all criminal cases in our data that faced a judge and received a conviction. The structured sentencing grid is defined by a convicted offense class and prior points group (see Appendix Figures B1–B2). Our stringency measure is a judge-level average of the residuals from this regression. To compute the change in county-level stringency between years $t - 1$ and t , we compute a judge \times county \times year specific average of the residuals that excludes any cases that the judge heard in the same county in year t or $t - 1$. This approach follows Chetty et al. (2014), and it ensures that year-to-year changes in county-level stringency are unrelated to changes in a judge’s tendency to impose active sentences. We exclude judge \times year pairs with fewer than 50 observations.

For Sections 4–5, we similarly begin by running a regression of an active sentence indicator on dummies for the structured sentencing grid. The sample for this regression all includes criminal defendants that faced a judge and received a conviction, but in this case we include only defendants who are *not* linked to any student in our education data (i.e., a leave-out sample). Our stringency measure is the judge-level average of the residuals from this regression. We exclude judges with fewer than 50 observations in the leave-out sample.

- **Math score.** Math scores include scores on both end-of-grade 3–8 math exams and the end-of-course high school Algebra exam, which is most commonly taken in 9th grade. We standardize these scores to be mean zero and standard deviation one in the full population of test takers in each year. If the student took an exam multiple times in the same academic year, we use their first score. If the student took an end-of-grade and the Algebra exam in the same year, we use the average of their two standardized scores.
- **Neighborhood SES percentile.** We follow Norris et al. (2021) in defining a measure of mean socioeconomic status (SES) in the neighborhood where the student lives. We define neigh-

neighborhoods using the student's Census Block Group as reported in the GEOADDRS 20**PUB datasets from NCERDC (see Appendix C.2.2). Since NCERDC collects addresses at the start of the academic year, we define our neighborhood SES measure based on the student's address in the *next* academic year; this allows neighborhood quality to potentially change *in the disposition year* in response to an incarceration.

To define neighborhood SES, we use the B17017 5-year average American Community Survey dataset for the years 2011–2015 (available at: <https://data.census.gov/cedsci/>). For each Census Block Group, we compute the fraction of households with income above the poverty level in the past 12 months. We then rank all Census Block Groups in North Carolina based on this measure, and use percentile rank as our outcome variable. The value one represents the neighborhood with the lowest poverty rate, and zero represents the neighborhood with the highest poverty rate.

- **Number of suspension days.** The sum of the student's in-school and out-of-school suspension days during the academic year.
- **Offense.** Unless otherwise noted, we define a defendant's offense at the type \times class level. Offense type is either felony, misdemeanor, traffic, or clerk to decide. Felony offense classes are A, B1, B2, C, D, E, F, G, H, I, or clerk to decide. Misdemeanor/traffic offense are A1, 1, 2, 3, or clerk to decide. Offense type \times class is the relevant level for North Carolina's structured sentencing regulations (see Appendix Figures B1–B2). Throughout our analyses, we exclude Class 2 and Class 3 traffic offenses (mainly speeding offenses), as these cases almost never result in an active sentence. In some robustness analyses, we define offenses at the level of a 4-digit offense code (e.g., Misdemeanor Larceny).
- **Student has criminal charge.** An indicator equal to 1 if the student *themselves* has a criminal charge in a given year based on our merge of the court (ACIS) and education (NCERDC) datasets (see Appendix C.4). The ACIS data includes each defendant's exact date of birth, while the NCERDC data includes each student's birth month and year. Thus we define the student as having a criminal charge if the defendant that we linked to the student has the same birth month and year as the student. This outcome is defined only for the 2009–2014

academic years because we rarely observe criminal cases with offenses that occurred before before July 1, 2008 or after June 30, 2014. Note that we do not observe criminal activity after the student has left K–12 public school because our merge is based on addresses in the NCERDC data.

C.2 Data

Our analysis uses administrative datasets from two main sources: 1) the North Carolina court system; and 2) the North Carolina Education Research Data Center (NCERDC). For supplementary analyses we also use data from the North Carolina Department of Corrections. We describe each of these datasets below.

C.2.1 Court data

Our first data source is from the North Carolina court’s Automated Criminal/Infractions System (ACIS). These records cover all court cases in the state with information on the defendants, offenses, and sentencing outcomes. Our extract from this system includes all cases in which the date of last update was between July 1, 2009 and June 30, 2014. Our analysis focuses on criminal cases, which include felonies, misdemeanors, and criminal traffic offenses such as DWIs and DWLRs (Driving While License Revoked). We also have data on non-criminal infractions during this time period. We included infractions in the merge with the education records as described below, but we exclude infractions from our analyses because these cases do not result in active sentences.

Most of the variables in our analysis come from the ACIS’s two main datasets:

1. *Case Records (CRCASES)*. This dataset includes an observation for each criminal case and includes characteristics of both defendants and their cases. We observe each defendant’s race, gender, and age, as well as their exact address at the date of last update in the system. Case characteristics include the date of origination and the court county and type (district/superior).
2. *Offense Records (CROFFNS)*. This dataset includes an observation for each charge of each criminal case, with information on the offense and disposition outcome. These records in-

clude the date of arrest, arraignment, and disposition, as well as a separate 4-digit offense code for each of these three events.³⁵ We observe the defendant's plea, verdict, and type of disposition (e.g., judge, jury, or offense dismissed). Importantly for our empirical strategy, the dataset includes the judge's initials if a judge was involved in the disposition. Lastly, this dataset also includes information on the sentencing outcome, including the type of sentence (active, intermediate, community), minimum/maximum sentence lengths, structured sentencing offense class, defendant's prior points, probation, and fines/court fees.

We also define a few variables based on the datasets *JATABLE*, *JTABLE*, *CRSPCOND*, which provide details on sentencing outcomes such as split sentences and credit for time served.

We merge the ACIS datasets using the case identifier (*crrkey*) and charge line (*cro1no*). Below we describe how we merge the court and education records and define the sample of criminal defendants for our analysis.

C.2.2 Education data

Our education data were provided by the North Carolina Education Research Data Center (NCERDC). The NCERDC contains longitudinal information on all public school students in North Carolina from kindergarten through the end of high school. Unless noted below, we use records that cover the 2006–2017 academic years, where 2006 represents the academic year from Fall 2005 to Spring 2006.

Our variables come from the following NCERDC datasets:

1. *Student Demographic and Attendance Data (ACCDEMOPUB20**)*. These datasets are at the student/semester level and cover all grade levels from K–12. They contain information on the schools students attended during the year, their grades, and days of absence. They also include demographic characteristics such as age, race, and gender. We use these datasets to define the full population of students in public school in a given year.
2. *Current Test File (CURTEST_PUB20**)*. These datasets include information on student exam performance with a separate observation for each student/year/test. We use the grade 3–8

³⁵ Offense codes are further aggregated into types (Felony, Misdemeanor, and Traffic) and classes (e.g., Felony F, Misdemeanor 2).

end-of-grade reading and math scores, and the end-of-course Algebra and English I exams that students take in high school. We use only 2008–2017 test scores from these datasets, and we keep a student’s first attempt at each test in a given year in cases where students repeat exams. Our main outcome variables are scale scores normalized to have mean zero and standard deviation one within the population of all students who took the same test in the same year.

3. *Student-Level Academic Summary (MB_20**_PUB/PCAUDIT_PUB20**)*. These datasets include both demographic and test score variables at the student/year level for grade 3–12 students. We use these datasets to measure end-of-grade reading and math scores in 2006–2007, which are not systematically available in the Current Test Files. We also use these datasets to measure economic disadvantage and to fill in missing values of days of absence.
4. *Student Offense-Consequence Data (MASTSUSP20**)*. These datasets include an observation for each disciplinary incident that schools report to NCERDC. They contain information on the incident date, type (e.g., fighting, truancy), and consequences (e.g., in- or out-of-school suspension). We use only disciplinary incidents in 2008–2017, as the data coverage is less in 2006–2007.
5. *Geocoded Student Address Information (GEO_REF_95_09/GEOADDRS_20**PUB)*. These datasets include students’ geocoded addresses measured at the beginning of each academic year. We use the geocoded address to define student moves, and for the merge with the ACIS data as described below. These datasets also include information on the Census blocks or tract of each address, which we use to merge in the following geographic variables:
 - 5-year average Census tract median income measured in 2010 from the American Community Survey.³⁶
 - Latitude and longitude of 2010 Census blocks.³⁷
 - Data on child incarceration outcomes by Census tract from the Opportunity Insights project.³⁸

³⁶ Obtained in November 2020 from: <https://data.census.gov/cedsci/>.

³⁷ Obtained in February 2021 from: <https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2010.html>.

³⁸ Obtained in February 2021 from: <https://opportunityinsights.org/data/>.

- A crosswalk between 2000 and 2010 Census blocks.³⁹

6. *Dropout Data — Student Records (MASTDROP20**)*. These datasets contain information on students who are reported as dropping out from school, including the date and reason for dropout.

7. *Transfer, Dropout, and Graduation Data (EXIT_PUB20**)*. These datasets contain information on students who are reported dropping out, transferring, or graduating from high school. We use these datasets to measure each student’s year of high school graduation, if any.

We collapse each of these datasets to the student/year level by computing, for example, a student’s first attempted test score or whether they had any suspension during the year. We then merge the collapsed datasets using the student identifier (*mastid*) and academic year. Below we describe how we merge the education and court records and define the sample of students for our analysis.

C.2.3 Prison data

Our final dataset includes prison records from the North Carolina Department of Corrections (DOC). These records cover all individuals with active prison sentences or probation managed by the DOC from the 1970s up through 2016. We use this data to measure the incidence and timing of any prison spells that result from a defendant’s case. Importantly, the DOC records do not cover North Carolina’s system of county jails. Many defendants who receive short sentences serve their time in county jails, which we cannot observe in our data.

Our main dataset from the prison records is the *Sentence Components (OFNT3CE1)*. This dataset contains information on each component of each prison sentence as committed by the court system. These records cover court cases ranging as far back as the 1970s up through 2016. The variables include the case number (*cmcaseno*) and county (*cmcocnvt*), which we use to reproduce the case identifier in the court data (*crrkey*) to merge the court and prison records. Our main outcome from this dataset is an indicator for having any active prison sentence resulting from a given case, which is defined by commitment prefixes (*cmprefix*) that begin with letters.⁴⁰

³⁹ Obtained in November 2020 from: <https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2010-census-block-record-layout.html>.

⁴⁰ Probation sentences have commitment prefixes that begin with numbers.

We note that the DOC records only cover North Carolina’s prisons, and not its system of county jails. Many criminal defendants who receive active sentences—particularly short sentences—serve time in county jails, which we cannot observe in our data.

C.3 Court data cleaning

This section describes two important data cleaning steps that are necessary to prepare the ACIS court data for our analysis.

C.3.1 Creating a defendant panel

In the ACIS data, the most granular unit of observation is a charge (`cro1no`) associated with a particular case number (`crrkey`) (see Appendix C.2.1). Many cases contain multiple charges, and it is common for multiple case numbers to refer to the same criminal event. Thus an important step in data cleaning is to combine charges and cases that are associated with the same criminal activity.

We do this by converting the court data into a defendant panel that has one observation per defendant and disposition event. A disposition event is any update in the defendant’s case, including the addition or dismissal of charges, superseding indictments, and judge rulings. Disposition events are identified in the offense records (CROFFNS) by the date variable `crrddt`.

We create a defendant panel in three steps. First, we create a unique ID number for individual defendants.⁴¹ This identifier is based on the defendant identification variables that we observe in the court data: full name (`crrnam`), date of birth (`crrdob`), driver’s license number (`crrdln`), last four SSN digits (`crrssn`), and street address (defined by `crradd`, `crrcty`, `crrdst`, and `crrzip`). We assign the same ID number to observations that match exactly on full name and at least one of the other four identification variables. To allow for typos or variations in the recorded name, we also use a fuzzy match of names in combination with the requirement that observations match on at least two of the other identification variables.

Second, we use the defendant ID to collapse the court data to a dataset with one observation per defendant/disposition event. If there are multiple charges or case numbers associated with

⁴¹ Creating a defendant ID also allows us to examine defendant recidivism outcomes.

a defendant/disposition pair, we keep the observation with the most serious 4-digit offense code at the time of arrest (`croffc`). We define the most serious offense as the one that has the highest offense class.⁴² If there are multiple offenses with the same class, we break ties using the more common 4-digit offense code as computed in our data. Throughout the paper, when we refer to a defendant’s “offense” for a given disposition, we are referring to the most serious offense as defined by these criteria.

This most serious offense determines the case number (`crrkey`) and charge line (`crojno`) that we include in the defendant panel. This defines the defendant and case characteristics that we use in our analysis, including defendant demographics and case county/type. The most serious offense is almost always the observation that is associated with a judge ruling, if one exists. In rare instances, a judge may rule only on a less serious charge. In these instances, we recover the judge’s identity and verdict associated with the less serious charge, but we continue to use the most serious charge to define the defendant’s offense.

Third, we define outcome variables for our analysis that reflect all of a defendant’s charges at a given disposition event. Our main defendant outcome is an active sentence, which we define as an indicator for an active sentence resulting from *any* of the defendant’s charges at that disposition event. Similarly, we use the *maximum* value for other sentencing outcomes, e.g., the maximum sentence length, the maximum probation term, or any prison sentence.

C.3.2 Defining judges and courts

In the ACIS data, judges are identified by the judge code variable `crdjno`, which is available in the offense records (`CROFFNS`). This code is typically the judge’s two- or three-letter initials, although other values are sometimes used. In most cases, the judge code identifies a single judge throughout the five years of our data extract, but sometimes two or more judges have the same code. In addition, the dataset does not indicate whether the judge is a district or superior court judge, which is important for the implementation of our judge stringency design.

To identify unique judges and their type (district/superior), we exploit the fact that the ACIS data includes both the case county and case type at the time of filing.⁴³ Cases are almost always

⁴² Specifically, the offense class ordering is Felony A > ... > Felony I > Misdemeanor A1 > ... > Misdemeanor 3 > Traffic A1 > ... > Traffic 3 > Infraction.

⁴³ The case’s county is indicated by the first three digits of the case number `crrkey`. Case types are defined by the

heard in the same county as their initial filing. On the other hand, it is more common for superior or district court judges to hear cases that were initially filed as the opposite type; this can arise from changes in the severity of the defendant's charges or from scheduling constraints.

We first group case counties into both *districts* and superior court *divisions*. In North Carolina, district court judges work in *districts*, which are comprised of either a single large county or a couple of smaller counties (see Figure B3). Superior court judges rotate to different courts in the same *division*, which are groups of *districts* (see Figure B4).

We then define judge types based the case types that are associated with each judge code (*crdjno*). We compute the total number of district court cases that are associated with a judge code/*division* pair across all years in our data. Similarly, we compute the total number of superior court cases for each judge code/*division* pair. We classify a judge code/*division* pair as a superior court judge if the number of superior court cases exceeds the number of district court cases. We classify the pair as a district court judge if the opposite is true.

Finally, we identify unique judges based on a threshold for the number of cases associated with each judge code in a given *district* or *division*. For district court judges, we assume that a judge code/*district* pair is a unique judge if either of the following conditions holds: 1) the pair is associated with 100 or more cases in any year; or 2) the pair is associated with 250 or more cases across all years. We use the same criteria for superior court judges, except we use judge code/*division* pairs.

We assign a unique ID number to each judge that we identify based on this procedure, and we use this number to define judges for all of our analyses. We identify 1,031 unique judges during our sample period, including 730 district court judges and 301 superior court judges. These judges are derived from 713 unique values of the judge code variable *crdjno*; judge codes that we assign to multiple ID numbers are often common two-letter initials (e.g., JH). Our main results are similar if we include only judge IDs that are associated with a single judge code value, which minimizes the potential for misclassification.

Throughout the paper, we define a "court" as a county/judge type pair. County is the location where a case is filed, and judge type (district/superior) is based on the above classification.

variable *crrtyp*, which appears in the case records dataset (see Section C.2.1).

C.4 Merge

To link our court and education records, we use a crosswalk between exact addresses in the two datasets that was created by the NCERDC. We sent the NCERDC a dataset with the address variables that are available in the court records, which include street address (*crradd*), city (*crrcty*), state (*crrdst*), and zip code (*crrzip*). The sample for this dataset included both criminal cases and infractions from July 1, 2009 through June 30, 2014, as described in Section C.2.1 above. The NCERDC linked these variables to the geocoded address identifier available in their records (*geo_addrid*) using their confidential information on exact address. We received back from the NCERDC a dataset that provides a crosswalk between *geo_addrid* in the education data and the address variables in the court data.⁴⁴

Appendix Table C1 reports summary statistics for addresses in the NCERDC and ACIS datasets and in the linked sample. Column (A) reports statistics for all student address identifiers that appear in the NCERDC data from 2006–2017. Column (B) includes all defendant addresses in the ACIS data from 2010–2014. Column (D) shows the subset of ACIS addresses that were linked to an NCERDC address, while column (C) includes unlinked addresses. Panel A shows that 48 percent of defendant address observations in the state of North Carolina were linked to an NCERDC address (4,443,522 out of 9,191,204). This is broadly comparable to the proportion of U.S. households that had children under the age of 18 in 2010 (roughly 45 percent), although these statistics can differ for many reasons.

Panels B–D of Appendix Table C1 provide evidence that NCERDC’s merge of addresses was high quality. The proportion of addresses that are in the largest counties (Panel B) and cities (Panel C) is broadly comparable across the datasets and samples, although these proportions also vary for reasons unrelated to merge quality (e.g., criminal activity and sorting of households with children). More convincingly, Panel D shows that the vast majority of defendant and student addresses in our linked sample have the same county and Census Block Group. To compute this statistic, we geocoded defendant addresses in the ACIS data to obtain the Census Block Group of each address, as this variable also appears in the NCERDC data.⁴⁵ Among addresses where these

⁴⁴ NCERDC does not collect address information from most charter schools. Thus while charter school students are included in many NCERDC datasets, nearly all students in our analyses samples attended traditional public schools.

⁴⁵ The notes to Appendix Table C1 provide details on our geocoding of defendant addresses.

TABLE C1: Merge statistics for NCERDC and ACIS addresses

	(A)	(B)	(C)	(D)
	All NCERDC addresses	ACIS court data		
		All ACIS addresses	Not linked to NCERDC data	Linked to NCERDC data
Panel A. Number of addresses				
# observations	19,023,642	9,660,294	5,216,747	4,443,547
# observations in North Carolina	19,023,642	9,191,204	4,747,682	4,443,522
# unique addresses	3,225,276	4,534,249	2,719,509	1,814,740
Panel B. Proportion of addresses by county				
Mecklenburg	0.075	0.107	0.136	0.083
Wake	0.078	0.086	0.067	0.102
Guilford	0.039	0.064	0.077	0.053
Cumberland	0.036	0.051	0.033	0.067
Forsyth	0.028	0.044	0.048	0.040
Durham	0.026	0.029	0.015	0.041
All other counties	0.717	0.618	0.623	0.614
Panel C. Proportion of addresses by city				
Charlotte	0.064	0.094	0.113	0.075
Raleigh	0.037	0.054	0.044	0.064
Greensboro	0.024	0.041	0.047	0.034
Fayetteville	0.028	0.040	0.026	0.055
Winston Salem	0.020	0.035	0.038	0.032
Durham	0.025	0.028	0.016	0.041
All other cities	0.802	0.708	0.716	0.700
Panel D. Proportion of linked address with matching locations				
Prop. with same county				0.998
Prop. with same Census Block Group				0.968

Notes: This table displays summary statistics for our merge between addresses in the court (ACIS) and education (NCERDC) datasets. Column (A) includes all student addresses in the NCERDC records from 2006–2017. Column (B) includes all defendant addresses in the ACIS records from 2010–2014. Column (D) includes the subset of ACIS addresses that were linked to an address in the NCERDC data based on the merge process described in Appendix C.4, and column (C) includes unlinked addresses.

Panel A reports the number of address observations and the number of unique addresses in each sample. Panel B shows the proportion of addresses that are in each of the six largest counties in North Carolina. Panel C shows the proportion of addresses that are in each of the six largest cities in North Carolina. We define the largest counties/cities based on the number of criminal offenses in the court data (column (B)), and we exclude addresses outside of North Carolina in computing these proportions.

Panel D shows the proportion of the linked addresses in column (D) that have the same county and Census Block Group. Information on counties and Census Block Groups are available in the NCERDC data. To define these variables in the ACIS data, we geocoded defendant addresses using the `censusxy` package for R (available at <https://cran.r-project.org/web/packages/censusxy/vignettes/censusxy.html>). This geocoding used the ACIS variables on the defendant’s street address, city, state, and zip code. From this process, we obtained the county and Census Block Group of 73 percent of all defendant addresses in the ACIS data and 85 percent of all linked addresses. Panel D reports the proportion of linked addresses that have the same county and Census Block Group among the subset of addresses for which these variables are defined in both datasets.

variables are defined in both datasets, we find that 97 percent of addresses are in the same Census Block Group, and nearly all are in the same county.

Using the address crosswalk, we link students and defendants who live at the same address in the same year. Specifically, to attach a student to a defendant, we require that the student’s reported address in a given academic year matches the year of last update of the defendant’s case

in the court data. Addresses in the NCERDC are collected at the beginning of the academic year; for example, addresses for the 2013–2014 academic year were based on where students lived in mid-2013. These students are linked to defendants who have the same address and for which the date of last update is between January 1, 2013 and December 31, 2013.

There are two potential sources of mismatch in our merge process. First, we may link students and defendants who are not related if families move around the time of their involvement with the criminal justice system. Addresses in the court data are updated if defendants move after their initial arrest; however, these updates may not be systematic, or families may have moved at some point during the year between the date of last update and the start of the academic year. We have experimented with different time windows that we use to link defendants and students by, for instance, using defendant years defined from July 1, 2013 to June 30, 2014 in the above example. These alternate time windows have only a minor impact on the composition of our final sample given the additional sample restrictions discussed below, and thus they do not significantly affect our main findings.

A second and more important source of potential mismatch arises from apartment buildings and other multi-unit addresses. Most addresses in the court data do include information on the unit number if it exists, but the geocoded addresses in the NCERDC data (`geo_addr_id`) do not distinguish between different units in the same building. This causes us to link to students to unrelated defendants in multi-unit buildings, which include apartment complexes, trailer parks, and homeless shelters. To isolate residences where the student and defendant are likely to be related, we impose restrictions on the number of students who are matched to each defendant, and the number of defendants who are matched to each student. We discuss these sample restrictions in Appendix C.5.

After we have attached a student to a defendant using the above process, we link the student to all of the defendant's cases that we observe in the court data. This step uses the defendant ID number described in Appendix C.3.1. This allows us to identify the most serious court event that a student was exposed to during our sample window, which we use to define our direct exposure samples as discussed below. Since moving is a common event for both students and defendants, the benefits of linking students to the full set of relevant cases are likely to outweigh the costs of any additional mismatch.

As further evidence on the quality of our merge, Appendix Table C2 shows that there is a strong relationship between the race/ethnicity of defendants and students in our linked sample. Column (A) reports the defendant’s race/ethnicity using the categorization in the ACIS data. Columns (C)–(I) report the proportion of linked students in each race/ethnicity group based on the NCERDC categorization. Panel A includes all linked defendants/students (as in column (C) of Table 1). The proportion of students who have the same race/ethnicity as the defendant ranges from 67–82 percent in each category, as indicated by the bold numbers on the diagonal. If we restrict to only the five categories that appear in both datasets (White, Black, Hispanic, Asian, and Native American), we find that 84 percent of defendants and students have the same race/ethnicity (column (J)). Panel B shows that the proportion with matching race/ethnicity rises to 87 percent in our direct exposure event study sample, which restricts to defendants and students who are more likely to be related (see Appendix C.5).

C.5 Samples

Our paper uses five different samples to examine the impacts of community-level exposure to incarceration and its mechanisms. In Section 3, we use all North Carolina public school students for our judge turnover strategy (column (A) of Table 1). In Section 4, we use two samples that are appropriate for our event study and judge stringency strategies to examine the direct effects of exposure to household incarceration (columns (A)–(B) of Appendix Table A5). In Section 5, we examine the indirect effects of incarceration by defining two samples of students who were classmates of children in our direct exposure samples (columns (D)–(E) of Appendix Table A5). We provide details on each of these samples below.

C.5.1 Judge turnover sample

For Section 3, our judge turnover sample includes all students who attended a North Carolina public school in 2010–2014. This includes any student who appears in the Student Demographic and Attendance Data from NCERDC (see Appendix C.2.2) for the 2010–2014 academic years. Many of our regressions restrict to students who have Math or English scores as defined in Appendix C.1. In some regressions, we restrict to the subset of students who were *not* linked to the

TABLE C2: Comparison of defendant & student race/ethnicity in linked samples

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
Proportion of students by race/ethnicity									
Defendant race/ethnicity	# linked students	White	Black	Hispanic	Asian	Native Amer.	Multi-race	Pacific Island.	Prop. with same race in cols (C)–(G)
Panel A. All linked students									
White	430,053	0.814	0.062	0.063	0.011	0.008	0.041	0.000	0.849
Black	320,375	0.066	0.802	0.071	0.010	0.005	0.046	0.000	0.841
Hispanic	93,491	0.067	0.073	0.821	0.009	0.004	0.025	0.000	0.843
Asian	8,614	0.091	0.068	0.044	0.733	0.004	0.057	0.004	0.780
Native American	10,521	0.118	0.093	0.034	0.043	0.667	0.045	0.000	0.699
Other	21,429	0.179	0.107	0.386	0.234	0.010	0.081	0.003	
(Missing)	7,281	0.362	0.340	0.177	0.042	0.020	0.056	0.002	
Total	891,764	0.433	0.333	0.154	0.023	0.014	0.043	0.001	0.843
Panel B. Students in direct exposure event study sample									
White	61,149	0.832	0.061	0.046	0.006	0.010	0.045	0.000	0.871
Black	53,781	0.059	0.847	0.037	0.005	0.006	0.046	0.000	0.888
Hispanic	8,587	0.083	0.073	0.791	0.007	0.005	0.040	0.000	0.825
Asian	688	0.068	0.080	0.019	0.762	0.009	0.062		0.812
Native American	2,064	0.120	0.072	0.023	0.013	0.727	0.046		0.761
Other	1,327	0.181	0.122	0.442	0.136	0.016	0.103		
(Missing)	1,233	0.405	0.280	0.168	0.050	0.031	0.067		
Total	128,829	0.433	0.393	0.096	0.011	0.020	0.046	0.000	0.873

Notes: This table compares the race/ethnicity of defendants and students who we linked based on the merge process described in Appendix C.4. Panel A includes all linked defendants and students, which is the sample reported in column (C) of Table 1. Panel B includes students in our direct exposure event study sample, which is the sample reported in column (A) of Appendix Table A5. If the student was linked to multiple defendants, we use the defendant with the most serious offense as defined in Appendix C.5.

Column (A) reports the defendant’s race/ethnicity using the categorization in the ACIS court data. Column (B) reports the number of linked students. Columns (C)–(I) report the student’s race/ethnicity using the categorization in the NCERDC education data. The numbers in these columns represent the proportion of students in each row with a given race/ethnicity. Column (J) reports the proportion of defendants and students who have the same race/ethnicity. For this column, we use only observations in which the defendant and student *both* report a race/ethnicity in one of the following five groups: white, Black, Hispanic, Asian, and Native American.

court records using the merge procedure described in Appendix C.4 (column (B) of Table 1).

C.5.2 Direct exposure samples

For Section 4, we define two samples of students who were directly exposed to a household incarceration: one for our event study strategy, and another for our judge stringency strategy. These samples differ because of the different sources of identification. Nonetheless, the students in these two samples have very similar characteristics (columns (A)–(B) of Appendix Table A5), and there is significant overlap between the samples.

Event study sample

In defining our event study sample, our approach is to create a sample in which all students have a household member who was convicted of their most serious offense. Our event studies then compare students whose relatives received active sentences to those whose relatives were convicted of similar offenses but received non-active sentences.

To create this sample, we first identify the most serious court event for each defendant. The most serious event is an active sentence if it exists, and if not, it is a guilty verdict. If defendants have multiple active sentences or guilty verdicts, we break ties using the case with the highest offense class at arrest, and finally, the case with the earliest disposition.⁴⁶ We exclude defendants that never received a guilty verdict (e.g., if their case was dismissed or if they were found not guilty). We also exclude defendants whose most serious event involved a Class 2 or Class 3 traffic offense (mainly speeding offenses), as these charges almost never result in an active sentence.

Next, we restrict to students and defendants who are likely to be related, and we identify the defendant with the most serious court event among those that we linked to a given student. Our merge of students and defendants is at the address level (see Appendix C.4), which means that students who live in multi-unit buildings may not necessarily be related to the defendant. To improve the quality of matches, we exclude any student that is linked to more than 3 defendants, and we exclude any defendant that is linked to more than 6 students.⁴⁷ If the student is linked to multiple defendants, our event study sample uses only the defendant with the most serious court event (defined by the same criteria as in the previous paragraph).

Finally, we impose two restrictions related to the timing of the defendant's case disposition. First, we exclude the small number of cases that we observe that were disposed prior to July 1, 2009 or after June 30, 2014. Second, we require that students appear in the NCERDC data two years prior to the disposition of their relative's case. This ensures that we observe outcomes in the pre-period of our event study.

The resulting sample includes 128,829 students linked to 86,854 defendants, as shown in col-

⁴⁶ If the defendant has multiple convictions at the same offense class, we break ties using the most common 4-digit offense code as computed from our data. This is same sort order that we use to create the defendant panel (see Appendix C.3.1).

⁴⁷ Since the court data we provided to NCERDC included both criminal offenses and infractions, this restriction includes defendants that received minor traffic tickets or other infractions.

umn (A) of Appendix Table A5. All defendants in the event study received a guilty verdict, and the charge offenses are roughly equally divided between felonies, misdemeanors, and Class A1 or 1 traffic offenses. Our “treatment” group for the event study analysis is the set of students whose relatives received active sentences, which comprise 29 percent of the sample.

Judge stringency sample

To define the judge stringency sample, we first identify the most serious *charge* for each defendant in our court data. This is similar to the process used for the event study sample, except we do not condition on the verdict or sentencing outcome. Specifically, we select the case with the highest offense class at arrest, and then the earliest case to break ties.⁴⁸ We exclude cases that never faced a judge or that never received *any* verdict; these are primarily cases that were dismissed by the district attorney or hearings for probation violations. As above, we also exclude cases where the most serious charge was a Class 2 or Class 3 traffic offense.

Next, we exclude cases that involve judges for whom we cannot compute a reasonably precise measure of stringency. Specifically, we drop judges with fewer than 50 observations in our leave-out sample of defendants who were not linked to any student (see Appendix C.1). This restriction also excludes judges that we cannot cleanly identify in the data (see Section C.3.2).

Our final set of sample restrictions are the same as for the event study sample. We exclude any student that is linked to more than 3 defendants, and we exclude any defendant that is linked to more than 6 students. If the student is linked to multiple defendants, we keep only the defendant with the most serious charge as defined above. We cases that were disposed prior to July 1, 2009 or after June 30, 2014. Finally, since we also make pre/post comparisons in our judge stringency analysis, we drop students that do not appear in the NCERDC data two years prior to their relative’s case disposition.

The resulting judge stringency sample includes 118,416 students linked to 79,765 defendants, as shown in column (B) of Appendix Table A5. This sample is slightly smaller than the event study sample; although it includes defendants who did not receive a guilty verdict (21 percent of cases), it excludes defendants for which we cannot compute the stringency of their judge. Relative

⁴⁸ As with the event study sample, we use the most common 4-digit offense code if the defendant has multiple charges with the same offense class.

to the event study sample, the judge stringency sample includes a lower proportion of cases with active sentences (14 percent vs. 29 percent), but the characteristics of the students and defendants are otherwise very similar.

C.5.3 Indirect exposure samples

For Section 5, we define two samples of students who were indirectly exposed to a household incarceration. Our indirect event study sample includes students who attended the same school and grade as a child in our direct event study sample in the year of the defendant’s case disposition. Our indirect judge stringency sample includes students who attended the same school and grade as a child in our direct judge stringency sample in the year of the defendant’s case disposition. Throughout the paper we refer to indirectly-exposed children as “classmates” of the directly-exposed child since these students typically move through the school system together, and thus often share classes.

The two indirect exposure samples are very similar (columns (D)–(E) of Appendix Table A5) since indirect exposure to incarceration is common (see Appendix Figure A4). Our regressions in Section 5 are at the directly impacted student \times classmate \times year level, so classmates often appear in the regressions multiple times because they are linked to multiple students in the direct exposure samples.