

# Community Impacts of Mass Incarceration\*

Arpit Gupta<sup>†</sup>

NYU Stern

Christopher Hansman<sup>‡</sup>

Imperial College London

Evan Riehl<sup>§</sup>

Cornell University

February 14, 2022

## Abstract

Student achievement is lower in communities with high incarceration rates. To investigate this relationship, we assemble a new dataset which matches students' academic performance to the court records of household members. Exploiting the exogenous turnover of judges (who vary in their tendency to incarcerate), we find that increases in the incarceration rate negatively impact the test scores of children in the community, including those who do not directly experience the incarceration of a household member. To explain this result, we show that (i) direct household exposure to incarceration adversely impacts student test scores and behavioral outcomes, and (ii) these effects spill over onto the academic performance of classmates. While smaller in magnitude per student, spillover effects aggregate to explain the majority of the causal relationship between community level incarceration and test scores. Our results highlight important consequences of incarceration for access to opportunity within entire communities.

JEL codes: K14, K42, I24, J13

---

\*We thank Kara Bonneau, Ryan Fackler, Ethan Frenchman, David Hoke, John Rubin, and Jessica Smith for helpful conversations. We 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 their 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).

<sup>†</sup>Department of Finance, NYU Stern School of Business; Email: [arpit.gupta@stern.nyu.edu](mailto:arpit.gupta@stern.nyu.edu); Web: [arpitgupta.info](http://arpitgupta.info).

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

<sup>§</sup>Department of Economics and ILR School, Cornell University; Email: [eriehl@cornell.edu](mailto:eriehl@cornell.edu); Tel: +1 (607) 255-0395; Web: <http://riechl.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 some of the most underprivileged communities bear the brunt of this expansion. A typical child in Charlotte, North Carolina interacts with more than 50 schoolmates with incarcerated family members by the time they finish the 4th grade. Mass incarceration disrupts social networks and erodes public trust, which may explain why places like Charlotte have low rates of intergenerational mobility (Chetty, Hendren, Kline, and Saez, 2014; Chetty, Friedman, Hendren, Jones, and Porter, 2018). However, prior work in economics has focused primarily on how incarceration affects offenders and their family members, with less attention paid to impacts on the wider community. This gap in the literature reflects a fundamental challenge in credibly separating local incarceration rates from criminal activity, household sorting, and other aspects of schools and neighborhoods that shape children’s opportunities.

This paper focuses on the large negative association between student achievement and incarceration rates at the community level. Figure 1 highlights this relationship, plotting math test scores against incarceration rates at the census tract or school level in North Carolina. Our key contribution is to show that a meaningful fraction of the gradient—upwards of 25 percent—is causal. We do so using a judicial turnover design that exploits the year-to-year introductions and departures of judges who vary in their tendency to incarcerate convicted defendants in North Carolina’s county-level courts. We complement this approach with a series of student-level analyses focusing on the children of criminal defendants and their classmates to better understand the mechanisms driving our result.

We show that the bulk of the causal relationship between incarceration and achievement can be accounted for by the classmates of children who directly experience the incarceration of a family member. Directly impacted children experience adverse consequences on academic performance themselves, but make up a relatively small portion of the total student population. These children also increase their misbehavior, which spills over and adversely impacts the academic performance of their classmates through a behavioral disruption channel. Modest impacts on a large number of classmates aggregate to explain a meaningful fraction of the overall effect. Our results suggest that mass incarceration has broad negative consequences for local access to opportunity.

Assessing the community level impacts of incarceration is made possible through our creation of a new dataset that links criminal justice and educational records for the entire state of North Carolina. This linkage, and detailed information on the classrooms, schools, and neighborhoods of each student, allow us to examine the educational consequences for students directly exposed to incarceration, as well as spillovers onto the children that they interact with. In addition, the merge creates a comprehensive dataset in the United States. While past research has considered comprehensive defendant-student datasets in Sweden and Norway, or more limited regional linkages in the United States, our state-level coverage gives us the power to detect small but economically meaningful spillover effects.

To conduct our analysis, we begin with a judge turnover strategy to estimate the causal impact of county-level incarceration rates on test scores. Our approach is inspired by the teacher arrivals and departures strategy in Chetty, Friedman, and Rockoff (2014). We focus on year-to-year changes in average county-level *stringency* resulting from the entry and exit of judges who vary in their tendency to incarcerate criminal defendants.<sup>1</sup> We find that a one standard deviation increase in county-level stringency leads to a 15–20 percent increase in incarcerations, but is unrelated to local trends in criminal activity, demographic shifts, or prior test score changes. Our main finding is that these increases in community incarceration rates in turn lead to meaningful reductions in student test scores. A one standard deviation increase in county-level stringency generates a 2.5–3.5 percent of a standard deviation reduction in math scores, and a 1.5–2.0 percent of a standard deviation reduction in English scores. The effects are persistent, growing in magnitude over the course of at least three years, and are similar even when focusing on children with no direct exposure to the criminal justice system, pointing to broader community impacts of incarceration.

We then examine the mechanisms driving the relationship between incarceration and community achievement with two complementary empirical strategies that use linked student-defendant data. The first is an event study approach that looks at changes in achievement before and after students are exposed to an incarceration. To account for the impacts of of criminal activity and other confounds, we compare these children to a control group exposed to criminal defendants that were convicted of similar crimes but received probation or other lesser sentences. The second is a within-court judge randomization strategy that uses variation driven by the judge a

---

<sup>1</sup>Stringency here refers to the average rate at which a county incarcerates criminal defendants.

defendant sees, conditional on facing a particular type of charge. This strategy focuses on plausibly exogenous variation in incarceration driven by (i) persistent differences in incarceration rates across judges (as in the judicial turnover design) and (ii) quasi-random assignment of defendants to judges within counties. Both the event study and judicial randomization designs perform well in tests for pre-trends and balance on observables.

We find adverse impacts on test scores in math and reading for directly impacted students—those who share a household with an incarcerated individual—using these two identification approaches. Our event study estimates suggest that the incarceration of a household member reduces math scores by 1.5 percent of a standard deviation and leads to a quarter-day increase in suspensions. Test score effects are larger for Black students and are also larger when the incarcerated individual appears to be the student’s mother, consistent with a role for parental inputs. We also find negative consequences for a variety of behavioral outcomes (including absences, suspensions, and fighting incidents).

These increases in misbehavior for directly impacted students appear to disrupt the academic performance of their classmates. We repeat our judge randomization and event study strategies considering outcomes for students who are not directly exposed to an incarceration in the household, but who share a classroom with a directly affected child. We find meaningful reductions in test scores for these indirectly impacted classmates. For example, our event studies suggest that the direct exposure of one student reduces classmates’ math and English scores by 0.4 and 0.3 percent of a standard deviation, respectively.

While spillover effects are smaller than the direct effects of incarceration, they play a significant role in explaining the aggregate gradient between incarceration and test scores. Directly impacted children make up a relatively minor fraction of the population, and hence explain a very small part of the overall relationship. Alternatively, a substantial fraction of children are indirectly impacted. We estimate that causal effects on classmates are responsible for 6–9 percent of the unconditional aggregate gradient (and roughly 15 percent of the gradient, conditional on observables). As classmates make up only a portion of a student’s peer relationship—they interact with children in other grades and in a variety of non-academic settings—we consider these estimates to be a lower bound on the causal relationship explained by spillovers.

Our results do not necessarily imply that incarceration is an unmitigated negative policy. Our

findings are specific to the criminal justice system in North Carolina—which, like many in the United States, may do an inadequate job at rehabilitation. Other approaches around the world, for example, the well-studied systems in Scandinavia that emphasize reintegration, may lead to more positive outcomes for criminal defendants and their family members. 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 students' achievement and access to opportunity. Policymakers must weigh both the costs and benefits of incarceration, taking into account broader spillovers within communities.

Our paper relates most closely to interdisciplinary work on the relationship between increased imprisonment and a variety of community outcomes. This literature has emphasized the disproportionate local impacts that result from the spatially concentrated nature of incarceration, with lasting consequences along multiple dimensions. This includes the impact of parental incarceration on children's behavioral outcomes, such as aggression, which have the potential to affect the achievement of students who share the same classroom or neighborhood.<sup>2</sup> Incarceration shocks 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 incarceration itself (Pattillo, Western, and Weiman, 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 towards family formation, authorities, and the government (Rose and Clear, 1998; Lynch and Sabol, 2004). Finally, the cumulative impacts of incarceration may further increase local criminal activity in the long-run (Clear, Rose, Waring, and Scully, 2003). We draw on this large literature and contribute by providing causal evidence on the existence of community level impacts, and by highlighting the role of spillovers within the classroom as a key mechanism.

A novel contribution of our 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).

---

<sup>2</sup> 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).

More generally, Billings, Deming, and Rockoff (2014) find that children have lower achievement when they attend schools with more minority students. Criminal activity, arrests, and family demographics are each strongly correlated with local incarceration rates, making it difficult to isolate the role of sentencing decisions themselves. While these other factors are surely important, our findings suggest that there are widespread consequences of incarceration for access to opportunity within communities.

We also connect 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 outcomes. Our work is closely related to recent work which find mixed evidence on the effects of parental incarceration using judge randomization designs, primarily in contexts outside the United States. 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 outcomes in Colombia. Closest to our paper in the United States, Norris, Pecenco, and Weaver (2021) find no significant impact on children's academic outcomes using data from a county in Ohio, and weak evidence of behavioral disruptions including absenteeism and repeated grades. This research motivates our effort to study the direct impacts of incarceration in United States—a country with distinctively high rates of incarceration—using a sample large enough to provide the power to detect the economically meaningful results we discuss above.

More fundamentally, however, our central contribution to this literature is to shift the focus away from the direct impacts on the children of incarcerated individuals and toward the broader consequences of incarceration for entire communities. While directly affected children may be most acutely impacted, a much larger fraction of students are exposed to the criminal justice system through indirect channels. Our community-level approach provides novel evidence on the aggregate impacts of incarceration on test scores. The resulting estimates are an order of magnitude too high to be explained through the direct channel alone. Instead, we emphasize the indirect channel: widespread spillovers onto classmates and peers within the community.

## 2 Data and the Community Achievement-Incarceration Gradient

Estimating the impact of incarceration on community education outcomes is a challenging problem which requires both detailed incarceration and educational records. Establishing the mechanisms linking incarceration and those outcomes, in particular, is difficult due to the granularity of data required, the need for linkages across multiple datasets, and the obstacle of finding credible sources of identifying variation. We address these challenges through a novel merge between administrative datasets in North Carolina. We focus on this state because it provides extensive administrative data on courts, prisons, and the school system; it is a substantially sized state in the United States; and it is broadly representative of communities around the country in terms of demographic composition and exposure to incarceration.

### 2.1 Data Sources

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

**Court Records (ACIS).** Our first data source is the North Carolina Court’s Automated Criminal/Infractions System (ACIS). These records cover all criminal cases in the state in 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. We observe defendant demographics and characteristics of the case at origination such as its date, county, and court type (district or superior). The data contain detailed information on the criminal activity, including offense codes at the times of both arrest and conviction. Importantly for our strategy, the data 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, structured sentencing offense class, and defendant’s prior points. Throughout the paper we exclude lower-level traffic offenses (Classes 2–3), which rarely result in incarceration. Appendix C.2.1 provides details on the ACIS data, and Appendix C.3 describes our cleaning process for this data.

A key variable in this data is the outcome of incarceration. A central choice that 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: incarceration

in prison or jail. Because an active sentence necessarily entails incarceration, we use these terms interchangeably in contrast with community and intermediate sentences. 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 system and structured sentencing.<sup>3</sup>

**Education Records (NCERDC).** Second, we use longitudinal records from the North Carolina Education Research Data Center (NCERDC) that cover all K–12 public school students in the state from 2006–2017.<sup>4</sup> 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 incidents of fighting. The NCERDC data also include students' geocoded addresses each year, which facilitates our merge with the court data.<sup>5</sup> Appendix C.2.2 describes the NCERDC data in detail, and Appendix C.1 defines our key variables.

## 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. Specifically, we link students and criminal defendants who live in the same address in the same year, where addresses include street numbers, cities, and ZIP codes.<sup>6</sup> Appendix C.4 describes the details of the merge.

Our address-based merge has two important caveats. First, we do not observe the relationship

---

<sup>3</sup> 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 will compare 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 will generally bias us against finding an effect of incarceration on other outcomes.

<sup>4</sup> Throughout the paper, we define time by academic years rather than calendar years. For example, the 2010 academic year includes defendant and student outcomes measured from July 1, 2009 through June 30, 2010.

<sup>5</sup> The education data covers 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.

<sup>6</sup> After we have made a link between a student and 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.

between the student and the defendant. As a result, our measure of *directly impacted* students reflects 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 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 make an inference on 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 co-habiting. Since student addresses are recorded only once per year, false matches can occur if families move mid-year.<sup>7</sup> 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 our direct exposure samples (see Appendix C.5).

### 2.3 Analysis Samples and Summary Statistics

Our analysis is organized into three parts, each focusing on different samples. We begin with a judge turnover strategy that focuses on the universe of public school students in North Carolina. We next explore the direct impacts of incarceration using data on students with a household member that is charged with or convicted of a criminal offense. Finally, we consider indirect impacts using data on students who have a classmate that is directly impacted. Each analysis requires a different sample construction which we discuss in turn below, with summary statistics provided in Table 1. Appendix C.5 provides a full description of how we construct each sample.

**Judge Turnover Sample.** Column (A) of Table 1 shows the sample for our judge turnover analysis, which we use to establish the impacts of county-level incarceration rates on student achievement. This sample includes the universe of public school students in the 2010–2014 academic years. In some regressions, we restrict to students who were *not* linked to the court data. Column (B) reports summary statistics for students who were not linked, while column (C) includes students who were linked to a criminal defendant. 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. Relative to other public school

---

<sup>7</sup> The ACIS records update addresses if defendants move, but mismatch could also arise from lags in this updating.

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.

**Direct Exposure Samples.** Columns (D) and (E) of Table 1 show summary statistics for our direct exposure samples. When considering direct impacts, we use two complimentary empirical strategies which require slightly different samples. Our event study approach compares students with household members who were *convicted* of similar crimes but differ in whether or not they ultimately received an active sentence. To create this sample, we identify the most serious convicted offense for the defendant(s) linked to each student and exclude defendants who did not receive a guilty verdict.<sup>8</sup> Our judge stringency approach, by contrast, compares students with household members that were *charged* with similar offenses but who were assigned to different judges. This differs from the event study sample in that we do not condition on the conviction or sentencing outcome, which may be endogenous to the judge. We do exclude cases that never faced a judge, and we drop judges for whom we cannot compute a reasonably precise measure of stringency (see Section 4.3).<sup>9</sup>

The event study sample has a higher active sentence rate than the judge stringency sample (29 percent vs. 14 percent) since it explicitly selects cases on this outcome, but otherwise children and defendant characteristics are very similar in the two samples. Both samples are much smaller than the full matched sample in column (C); this reflects our strict criteria to isolate clean matches, our exclusion of Class 2–3 traffic offenses, and the fact that a large fraction of cases are dismissed or resolved without ever reaching a judge. Children in our direct exposure samples have lower test scores and higher rates of misbehavior compared to those in the full matched sample.

**Indirect Exposure Samples.** Columns (F) and (G) of Table 1 show summary statistics for our samples of students who were *indirectly impacted* by household incarceration. We define students as indirectly exposed if they attended the same school and grade as a directly-impacted child in the year of the defendant’s disposition. Throughout the paper, we refer to these children as “class-

---

<sup>8</sup> The most serious conviction is an active sentence if it exists, and if not, a guilty verdict. If defendants have multiple active sentences or guilty verdicts, we break ties using the case with the highest offense class, and finally, the earliest disposition. Active sentences in this context reflect incarceration after conviction, as opposed to intermediate or community punishments.

<sup>9</sup> Because we are unable to link some cases with a well-defined judge, the number of observations in the judge stringency sample is slightly lower than in the event study sample.

mates” of the directly-exposed child since they typically move through the school system together, and thus often share classes. To align with our analysis of direct impacts, we define indirect exposure samples separately for our event study and judicial stringency approaches. But as we emphasize below, indirect exposure to incarceration is common. Thus our two indirect samples are very similar, and both are comparable to the full population of North Carolina students.

## 2.4 Baseline Evidence of an Achievement-Incarceration Gradient

Figure 1 shows two versions of the central empirical relationship of our paper: the large negative correlation between community incarceration levels 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 2010–2014. In this panel, school exposure is the average number of students who experience a household incarceration in a typical year as computed from our full match between the ACIS and NCERDC datasets. Panel B shows a similar negative relationship between neighborhood-level incarceration events and math scores, where we define neighborhoods as the Census tracts in which defendants and students live.

The gradients reflect sizeable correlation between incarceration and test scores: Column (A) of Table 2 shows results from OLS regressions that capture the linear slopes in Figure 1. 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. In other words, community incarceration levels are a powerful predictor of student achievement. Academic performance is substantially lower in areas exposed to significant incarceration activity.

Of course, an important limitation in Figure 1 is that this raw correlation need not reflect a causal relationship. Families likely sort to different neighborhoods and schools based on income, employment opportunities, crime levels, and a variety of other factors associated with local incarceration rates. To begin to disentangle the causal relationship from these selection issues, columns

(B)–(C) of Table 2 present more saturated OLS specifications that include student demographics as well as county and year fixed effects. Accounting for demographics and fixed effects reduces the magnitude of the gradient by 50–60 percent. Still, even conditional on these controls, we observe a robust and economically meaningful negative relationship between incarceration and achievement. This relationship persists when we exclude children with an incarceration in their household (column (D)), and even when we focus on children who are never matched to the court data (column (E)). Thus the negative relationship between incarceration and academic outcomes is not strictly driven by children in families that are directly involved with the criminal justice system.

While these more saturated OLS regressions reinforce the possibility that a causal mechanism is at play, potential confounds remain. In the next section, we develop a strategy based on the turnover of judges to isolate the causal effect of community incarcerations on student achievement.

### 3 Estimating the Causal Effects of Community Incarceration with a Judicial Turnover Design

The ideal experiment to estimate the impacts of community level incarceration on children would be to exogenously reduce the fraction of criminal cases that result in incarceration for a set of randomly selected 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 using 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 key intuition is that when a more stringent judge—one with a greater tendency to impose active sentences—enters a county, aggregate incarceration levels rise. We examine turnover due to both arrivals and departures, stemming from a combination of judicial elections, mid-term leaves, and pre-scheduled judicial rotations across counties.

### 3.1 Implementing the Judicial Turnover Design

Our strategy exploits 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 et al. (2014)'s strategy of using teacher arrivals and departures to validate teacher value-added estimates, and we follow their empirical specification. Specifically, for county  $c$  and year  $t$  (and judge  $j$ ), we define the change in 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 a weight representing judge  $j$ 's caseload in year  $t$  or  $t-1$ ,  $\omega_{jct}$ . We describe each component in turn below.

**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 A1 and A2). 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 a judge that is present in a particular county or year with a large number of relatively serious offenses (within sentencing cell) is 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  $c$  in 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 jackknife leave-out-means, this approach is intended to capture a measure of the judge's tendency that is not directly influenced by the specific time and place. This mean is defined on the basis of judge  $j$ 's tendency to impose active sentences in other counties (or in the same county in other years). For our main estimate of  $\mu_{jct}^{-\{t-1,t\}}$ , we limit our sample to judge-year pairs with at least 50 observations.

**Caseload Weights  $\omega_{jct}$ .** In computing the change in county-level stringency between years  $t - 1$  and  $t$ ,  $\Delta S_{ct}$ , 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 judges hear in that county, as captured by the caseloads weights,  $\omega_{jct}$ .

We take two approaches to these weights. First, we use each judge's actual caseload in county  $c$  and year  $t$ , and we refer to this as our *actual caseload* measure. In this approach, we define  $\omega_{jct}$  to be the fraction of all criminal cases in county  $c$  and year  $t$  that were heard by judge  $j$ . Under this approach,  $\Delta S_{ct}$  captures changes in average stringency that arise from both the arrival and departure of judges in each 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.

To capture changes in stringency that occur only because of judges entering or exiting the county, we also consider a version of weights that fixes  $\omega_{jct}$  across years, and we refer to this as our *fixed caseloads* measure. In this approach,  $\omega_{jct}$  is equal to zero if judge  $j$  is not active in county  $c$  in year  $t$ , and otherwise it is equal to judge  $j$ 's average fraction of cases heard in county  $c$  across all active years in our sample.<sup>10</sup> With this 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), and (iii) rotations across counties. We explore these different sources of variation in more detail in subsection 3.6 below.

Appendix Figures A3 and A4 show the average stringency of district and superior court judges in each county in 2010, along with boundaries for district and superior courts. There is substantial variation in the average tendency of judges to impose an active sentence at the county level; the standard deviation of county stringency is 7 percentage points (in active sentence units), and the

---

<sup>10</sup> We consider judge  $j$  to be active in county  $c$  if he or she sees at least 20 cases.

change from the 10th-90th percentile is 20 percentage points.

### 3.2 Changes in County Level Stringency Induce Changes in Incarceration Rates

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

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

in which  $\Delta Y_{ct} \equiv Y_{ct} - Y_{c,t-1}$  and  $Y_{ct}$  measures the log number of active sentences in a given county and year.<sup>11</sup> This is comparable to the first-stage in an instrumental variables approach, although we present only reduced form effects of county stringency throughout our analysis.  $\gamma_t$  are year fixed effects, and standard errors are clustered at the county level. Across all specifications, we scale  $\Delta S_{ct}$  so that one unit corresponds to one standard deviation in the baseline distribution of county stringency (0.07). In Panel A, column (B) shows that a one standard deviation increase in county stringency is associated with a 0.208 log point increase in the number of active sentences using our actual caseloads measure. Column (C) shows an increase of 0.166 log points when using the fixed caseload measure. In other words, a one standard deviation increase in county level stringency generates a roughly 15–20 percent increase in the annual number of active sentences, which equates to approximately 100–125 incarcerations in the average county. The median number of judges in a county-year pair is nine, so shifts in judge composition have the potential to generate sizable impacts on the incarceration rates experienced in local communities.

### 3.3 Balance in the Judicial Turnover Design

We next address the central challenge to the 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 towards more criminal activity, or as county demographics change.

We provide evidence against this concern in Panels B–D of Table 3. These panels show esti-

---

<sup>11</sup> 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  $\Delta Y_{ct}$ . The one exception to this is when  $\Delta Y_{ct}$  is measured in logs (Panels A–B of Table 3), in which case we estimate unweighted regressions.

mates from Equation 2 using different dependent variables. In Panel B, the outcome variables,  $\Delta Y_{ct}$ , measure changes in the log total number of criminal cases in a county between years  $t - 1$  and  $t$ . We see little evidence that county stringency correlates with aggregate caseloads. With our active caseload measure (column (B)),  $\Delta S_{ct}$  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). We find similar results using our fixed caseload measure (column (C)), with the exception of a small and marginally-significant relationship with the total number of cases. Joint  $F$ -statistics across all criminal case totals are 0.3 for the actual caseloads measure and 1.5 for the fixed caseloads measure. This suggests that effects on the number of active sentences are driven by judges' propensities to opt for incarceration, rather than by changes in the number or type of charges.

We also find that changes in county stringency are not significantly related to changes in the average characteristics of criminal defendants or the mean demographics of children who attend school in those counties. Panel C shows that  $\Delta S_{ct}$  is unrelated to changes in the average county level race, age, gender or criminal history of defendants. Panel D shows that  $\Delta S_{ct}$  is not related to changes in the age, gender, or socioeconomic status of students; 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.

An important takeaway from Table 3 is that our county stringency measures allow us to separate community incarceration rates from aggregate criminal activity. This addresses an important potential confounder in the descriptive results in Section 2, as well as in many analyses of the community impacts of incarceration.

### 3.4 Increases in County Level Stringency Reduce Children's Academic Achievement

We now show the main results in our judge turnover design: increases in county-level stringency reduce the academic performance of children who live in those counties. Given the robust relationship between  $\Delta S_{ct}$  and incarceration rates, and the absence of any meaningful relationship with criminal case totals, defendant characteristics, and student demographics, we interpret this as evidence of adverse causal effects of community level incarceration on student achievement.

Panel A of Table 4 highlights the negative test score impacts of county-level incarceration. We repeat the specification in Equation 2 but define  $\Delta Y_{ct}$  as changes in county mean test scores in either math or English. Columns (B) and (C) show that a one standard deviation increase in county stringency results in a decline of 0.024 to 0.036 standard deviations for math scores, and a 0.014 to 0.020 standard deviation decrease for English scores. We observe larger impacts on math scores than on reading and English, and observe slightly higher adverse impacts using the fixed caseload measure rather than the actual caseload measure.

Columns (D) and (E) of Panel A show that this adverse impact is not limited to children whose families are directly involved with the criminal justice system. In these regressions, we repeat the specifications in columns (B) and (C), but limit the sample to students that are never linked to a defendant in the court records. The estimates for both math and English scores are similar—if anything, slightly larger—and remain statistically significant. The fact that we observe a large adverse impact on children who do not experience the incarceration of household members suggests that the aggregate impacts of incarceration must operate, at least in part, through community-level factors which are shared by many children through indirect ties. We explore these possible mechanisms in more detail in Section 5.

To show that changes in county-level test scores are not driven by compositional shifts in the student population, Panel B again follows Equation 2, but replaces  $\Delta Y_{ct}$  with the average *individual* change in test scores. That is, we take the set of students  $i$  with a test score in both  $t$  and  $t - 1$ , compute  $\Delta Y_{ict} = Y_{ict} - Y_{ict-1}$ , and then average  $\Delta Y_{ict}$  within county. For our actual caseload measure, the point estimates are virtually identical to those in Panel A (although no longer statistically significant, given higher standard errors). For the fixed caseload measure, the point estimates are marginally smaller but remain statistically significant. Thus the same students perform worse on math and English exams following increases in the stringency of judges who serve in their county.

Under the assumption that the impact of county level stringency operates *only* through the channel of incarceration, these estimates imply that a large fraction of the neighborhood-level gradient in Panel B of Figure 1 is due to a causal effect of exposure to incarceration. For example, our actual caseload estimate in Table 4 suggests that a one log point increase in active sentences leads to a 0.115 standard deviation reduction in county average math scores.<sup>12</sup> This represents

---

<sup>12</sup> This is derived from a simple Wald-style ratio of the impacts of county stringency on math scores ( $-0.024$ ) and

just over 40 percent of the unconditional gradient shown in Panel B of Table 2 (column (A)), and is nearly identical to the estimated gradient after controlling for observables (column (C)). A similar calculation suggests that the causal effect of exposure to incarceration is responsible for nearly 25 percent of the raw neighborhood-level gradient in English scores. These results indicate that a non-trivial fraction of the aggregate relationship between incarceration and achievement is causal, and not simply an artifact of selection.

### 3.5 Leads and Lags in the Relationship Between Stringency and Test Scores

Our results in Table 4 have a causal interpretation under the assumption that changes in county-level stringency are unrelated to other trends that impact children's achievement. To further validate this assumption, Table 5 shows that changes in stringency in year  $t$  are not related to test score changes in *prior* years. In addition, we show that the decline in test scores in year  $t$  persists in later years, suggesting that increases in community incarceration rates having lasting impacts on student achievement.

To investigate timing effects, we modify Equation 2 to consider various differences for the outcome variable  $\Delta Y_{ct}$ . To evaluate pre-trends in active sentencing and test scores, we consider the differences in these outcomes between  $t - 3$  and  $t - 1$  as well as  $t - 2$  and  $t - 1$ . To examine the persistence of changes in stringency at time  $t$ , we consider differences in outcomes between  $t - 1$  and  $t + 1$  as well as  $t - 1$  and  $t + 2$ . In all cases, the change in county-level stringency,  $\Delta S_{ct}$ , is defined as the difference between  $t$  and  $t - 1$ .

Columns (A) and (B) of Table 5 show that changes in judicial stringency at time  $t$  do not correlate with trends in active sentences in the pre-period. However, they do predict relatively stable changes in active sentences *after* time  $t$  at all horizons we consider. This holds for both our actual caseloads and fixed caseload measures. Similarly, the remaining columns show little relationship between judicial stringency and test score growth in years prior to  $t$ . Again, we see a strong relationship at time  $t$  and longer horizons. The math and English coefficients are larger in magnitude at times  $t + 1$  and  $t + 2$  than at time  $t$ . These results suggest that community-level exposure to

---

log number of active sentences (0.208). This calculation also assumes linearity in the impacts of county stringency. Appendix Table A1 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/neighborhoods with the highest exposure.

incarceration has persistent and accumulative effects on student achievement. Across all specifications, tests of joint significance for both pre-period differences are insignificant and tests of joint significance for all three post-period differences are significant.

### 3.6 Decomposing Sources of Variation in County Level Stringency

Our above results use identifying variation from all factors that affect judges' caseloads in a given county and year. In this section, we decompose this variation into different components to better understand the underlying sources of judicial turnover and examine the robustness of our results.

Panel A of Table 6 shows a variance decomposition of changes in county-level stringency. Our interest is in the variance of  $\Delta S_{ct}$  (rather than  $S_{ct}$ ) since this is the variation that drives our analysis. Column (A) shows that the variance of  $\Delta S_{ct}$  is 0.018<sup>2</sup> in our actual caseloads measure, while column (B) shows that the variance is 0.014<sup>2</sup> in our fixed caseload measure. In other words, 60 percent of the total variation in  $\Delta S_{ct}$  is driven by the introduction or departure of particular judges, while the remaining 40 percent is driven by changes in the caseloads of judges who were present in the county in both years  $t$  and  $t - 1$ .

Columns (C)–(E) of Panel A examine the different reasons for judge arrivals and departures. These columns decompose our fixed caseload measure into three components: 1) judge elections, which occur every two years with judges serving 4- or 8-year terms; 2) mid-term leaves due to retirement, resignation, or appointment to a higher court; and 3) judges moving to other counties, mainly from the prescribed rotation of superior court judges. For this we use the same caseload weights,  $\omega_{jct}$ , as in our fixed caseloads measure, but we compute  $\Delta S_{ct}$  using only judges who arrive/depart for each reason.<sup>13</sup> We find that most judge turnover is due to elections (18 percent of the total variation) and mid-term leaves (13 percent of the total variation). Since superior court judges hear few cases relative to district court judges, only 3 percent of the total variation in actual caseloads is driven by county rotations.

The remainder of Table 6 replicates our above analyses using each source of variation. These results are similar to those in Tables 3–5, except we define changes in county stringency,  $\Delta S_{ct}$ , using only the source of variation represented in each column. Panel B shows effects on log total active

---

<sup>13</sup> We identify election turnover by linking judges in the court data to biannual election results using judge initials, names, and districts. We define mid-term leaves as all other reasons that judges disappear from the court data. We define county rotations as judges who stay in the data but appear in other counties.

sentences and mean student test scores. Panel C reports  $F$ -statistics from balance tests for each stringency measure, which measure joint significance in its relationship with county-level criminal cases, defendant characteristics, student demographics, and prior-year test scores. Columns (A)–(B) in Table 6 replicate our main results from above for our actual and fixed caseload measures. Columns (C)–(E) present new results using only variation from elections, mid-term leaves, and judge rotations.

For the two dominant sources of variation in judge turnover—elections and mid-term leaves—we find negative effects on student test scores and limited evidence of imbalance with respect to defendant or student characteristics. This suggests that both the election of more stringent judges and the mid-term departure of less-stringent judges lead to increases in incarceration that adversely effect student test scores. The test score effects are largest in magnitude for mid-term leaves, although this variation also has a larger effect on the county incarceration rate. We do not see evidence of test score impacts stemming from judge rotations across counties, but standard errors are large since there is limited variation.

## 4 Mechanisms: Direct Household Exposure

To explain the effects of county-level incarceration rates on student achievement, we begin by examining direct channels involving children who experience the incarceration of a household member. These children face disruption to their home environment, separation from family members, changes in household financial stability, and other potential challenges. In this section, we show evidence that direct exposure adversely impacts children’s outcomes using event study and within-county judicial stringency strategies. The incarceration of a household member meaningfully reduces student test scores and also leads to increases in school absences and misbehavior. These behavioral outcomes, in turn, also affect the academic performance these children’s classmates, a channel we explore in Section 5.

### 4.1 Baseline: Cross-Sectional OLS Comparisons

As a baseline, we estimate simple OLS regressions that compare two groups of students: those who directly experience the incarceration of a household member, and those whose household

member is charged with a similar offense but is not incarcerated (because the case is dismissed, the defendant is found not guilty, the defendant is put on probation, or for any other reason). For this regression, we use our direct judge stringency sample (see subsection 2.3 and column (E) in Table 1). For student  $i$  in year  $t$ , we consider:

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

Here,  $y_{it}$  are student educational outcomes measured in or after the year of the defendant's 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.<sup>14</sup>  $\theta_{o\tau(i)}$  is a fixed effect for offense class at arrest ( $o$ )  $\times$  court ( $c$ )  $\times$  disposition year ( $\tau$ ), each of which are characteristics of the defendant's case linked to student  $i$ .  $\mathbf{X}_{it}$  is a vector of student and defendant covariates.<sup>15</sup>

The results in column (B) of Table 7 show that exposure to incarceration is associated with lower achievement and higher rates of absence and misbehavior. Students directly exposed to incarceration have math and English scores that are 3.4 and 1.4 percent of a standard deviation lower, respectively. They are absent 0.4 more days per year on average, receive 0.15 more suspension days, and are 0.3 percentage points more likely to be disciplined for a fighting incident.

Although Equation 3 includes much finer controls than are often possible with cross-sectional analyses of incarceration, there may yet be unobserved factors that cause a defendant to receive an active sentence that are related to student outcomes. In the subsections below we consider two empirical strategies to address this identification issue.

## 4.2 Event Study

### Empirical Specification

Our event study strategy uses two sources of variation to examine the impacts of direct exposure to household incarceration on children's outcomes. First, we compare children whose household

---

<sup>14</sup> Note that many of our variables, including Active Sentence $_i$ , are characteristics/outcomes of the *defendant* linked to student  $i$ , rather than student  $i$  themselves. We define our direct exposure sample so that each student maps to only one case and one defendant (subsection 2.3), and so we use  $i$  as the subscript for these variables to reduce notation.

<sup>15</sup> 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.

members were convicted for the same class of offense in the same year, but varied in whether or not they received an active sentence. Thus our “treatment” group includes students who were exposed to an active sentence in their household, while defendants linked to students in our “control” group received probation, fines, or other non-active punishments. Second, we measure *within-child* changes in outcomes from before to after the defendant’s disposition.<sup>16</sup> A causal interpretation of these results requires that the outcomes for treated and control students would have followed parallel trends in the absence of a household incarceration.

Formally, our event study specification considers outcome  $y_{it}$  for student  $i$  in calendar year  $t$ :

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

$\eta_i$  is an individual fixed effect that accounts for time-invariant differences across students.  $\lambda_{og\tau(i)t}$  is a granular set of offense class ( $o$ )  $\times$  academic cohort ( $g$ )  $\times$  disposition year ( $\tau$ )  $\times$  calendar year ( $t$ ) fixed effects.<sup>17</sup> Our variable of interest is the interaction between an indicator equal to one if student  $i$ ’s household member received an active sentence,  $\text{Active Sentence}_i$ , and an indicator for years in or after the case disposition,  $\mathbb{1}\{t \geq \tau(i)\}$ . Intuitively, the granular fixed effects restrict identification to students who are in the same cohort, and whose household members were convicted in the same year with similar offenses. Our coefficient of interest,  $\delta$ , is a weighted average of the effects of exposure to 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 \(2021\)](#)).

We estimate Equation 4 using our direct event study sample (see subsection 2.3 and column (D) in Table 1) considering student outcomes in all available years (2006–2017). We cluster standard errors at the defendant level.

---

<sup>16</sup> Appendix Figure A6 shows that 40 percent of defendants in our treatment group appear in prison records in the year of their disposition, and nearly 20 percent were still in prison three years later. Only a small fraction of defendants with non-active sentence appear in prison. As we do not observe incarceration records from county jails, this figure should be interpreted as a lower bound on the fraction who were ever incarcerated.

<sup>17</sup> 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.

## Results

The event study estimation yields evidence that direct exposure to incarceration negatively impacts children's academic and behavioral outcomes. Column (C) of Table 4 shows estimates of the  $\delta$  coefficients from Equation 4. In Panel A, we find negative and statistically significant effects on both math and English test scores; exposure to a household incarceration reduces these scores by 1.5 and 1.2 percent of a standard deviation, respectively. Panel B shows substantial and statistically significant increases in days absent, the likelihood of any suspension, and days suspended, as well as a positive and marginally significant increase in fighting. Direct exposure leads to roughly an additional half day absent per year and a quarter of a day suspended. Across outcomes, the magnitudes in our event study specifications are similar in magnitude to those in column (B), suggesting that selection issues in the cross-sectional comparison are minimal given our granular controls.

To examine the identification assumption of parallel trends, we break out these effects into years from incarceration in Figure 2. For this we estimate a modified version of Equation 4, in which we replace the  $\mathbb{1}\{t \geq \tau(i)\}$  term with dummies for years relative to the defendant's disposition (omitting the year prior to disposition).<sup>18</sup> We plot the coefficients on these interaction terms from three years before through three years after conviction. We see little evidence of differential pre-trends between treated and control students in the years leading up to conviction. Post conviction, there is evidence of gradually increasing adverse effects across outcomes, pointing to persistent negative impacts of household incarceration. The pattern is particularly striking for days of suspension, a key measure of serious behavioral issues at school.

---

<sup>18</sup> Specifically, Figure 2 plots  $\delta_l$  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 we fix  $l = -1$  as our 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; our education data ranges from 2006–2017, and disposition years range from 2010–2014. We plot only coefficients from  $l = -3$  to  $3$ , as estimates are noisy beyond this range.

## **Robustness**

In Panel A of Appendix Table A2, we provide evidence that our results are not driven by differences in the offense severity or criminal histories of defendants who receive active sentences. Column (A) repeats our main specification as a benchmark. To test for differential reductions in charges prior to sentencing, column (B) replaces offense class at conviction with offense class at arrest in our fixed effects. Column (C) includes a more granular definition of offense at conviction—a four digit offense code rather than offense class. Column (D) includes additional controls for the defendant’s prior points interacted with a linear time trend, which tests whether children whose household members differ in criminal histories may have divergent trends in outcomes. Our results are similar across all specifications, suggesting that our event study results are driven by the defendant’s incarceration rather than the factors that influence the sentencing outcome.

## **Heterogeneity**

We examine heterogeneity in student outcomes in Panel A of Appendix Figure A9, which breaks out the event study coefficients on test scores across student demographics. For reference, we show the main effect for all students in the initial row. The relationship between incarceration and academic performance is most notable among 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. We do not see meaningful difference according to the gender of the student, but we do see differences based on the inferred relationship between defendants and students. The effects are most negative for defendants who are likely to be mothers (female defendants who are 20–40 years older than the student), consistent with the crucial role of mothers in child development.

### **4.3 Judicial Stringency Strategy**

Our event study results rely on the standard assumption of parallel trends. Although we see little evidence of differential pre-trends for students exposed to active versus non-active sentences, these children may still be differently affected by the criminal activity that influences the sentencing outcome. Our second empirical strategy allows us to estimate causal effects of direct exposure to household incarceration without the parallel trends assumption. We do so by comparing the

outcomes of children linked to defendants who faced judges that differ in incarceration stringency.

## Empirical Specification

Our strategy addresses the key identification concern with the baseline OLS regressions in Equation 3 using variation in judges' tendencies to impose active sentences. For judge  $j$  we denote this measure  $\mu_j$  and again refer to it as *judge stringency*.<sup>19</sup> 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.<sup>20</sup>

Our primary approach is a reduced form specification in which we directly regress student outcomes on our measure of judge stringency:

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

The coefficient of interest,  $\gamma$ , measures the impact of a household member being assigned to a more stringent judge on student outcomes. We include the same set of fixed effects,  $\theta_{oc\tau(i)}$ , as in Equation 3; this restricts identification to defendants who are at risk of seeing the same judges, as defined by their offense class, court, and disposition year. We also include the same vector of defendant and student characteristics,  $\mathbf{X}_{it}$ . We estimate Equation 5 using our direct judge stringency sample (see subsection 2.3 and column (E) in Table 1) considering student outcomes measured in or after the year of the defendant's disposition. We cluster standard errors at the judge level.

Our main 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 pre-determined assignments of judges to courtrooms at the discretion of the senior judge in consultation with county clerks. The frequent rotation of judges, at both district and superior levels, across courtrooms specializing in

---

<sup>19</sup> The only difference between our measures of judge stringency in Section 3,  $\mu_{jct}^{-\{t-1,t\}}$ , and Section 4,  $\mu_j$ , 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.

<sup>20</sup> We exclude judges with fewer than 50 observations in the leave-out sample.

different categories of cases ensures that judges oversee a variety of offense types.<sup>21</sup>

We also present results from an IV specification that uses judge stringency as an instrument for an active sentence. The first stage regression in this specification is:

$$\text{Active Sentence}_i = \gamma \cdot \mu_{j(i)} + \theta_{oct(i)} + \mathbf{X}_{it}'\zeta + v_{it}. \quad (6)$$

The second stage for this IV specification is Equation 3. Under standard independence, exclusion, and monotonicity assumptions, our IV approach estimates the LATE for students with complier household members: those who would receive an active sentence with a relative stringent judge in our sample, but would not under more lenient judges. There is an active literature on the econometrics of judge stringency regressions that highlights potential concerns in the interpretation of IV coefficients stemming from violations of exclusion and monotonicity (Mueller-Smith, 2015; Frandsen, Lefgren, and Leslie, 2019) and heterogeneity in causal effects (Kolesár, 2013; Goldsmith-Pinkham, Hull, and Kolesár, 2021). For this reason, we focus on the reduced-form coefficients,  $\gamma$ , in our discussion below, but include IV estimates for comparison with related work.

### First Stage and Balance

There are meaningful differences in the tendency of judges to impose active sentences. Appendix Figure A7 shows a histogram of judge stringency (Panel A); the standard deviation of  $\mu_j$  is 0.038, and the range from the 5<sup>th</sup> to the 95<sup>th</sup> percentile is 0.115. Stringency is also persistent for individual judges. In Appendix Figure A5, we compute stringency at the judge  $\times$  year level and examine its correlation across years. We find a correlation of 0.75 between judicial stringency in adjacent years, and the correlation remains high, at 0.52, even four years apart. The fact that sentencing differences persist over time is consistent with the interpretation of stringency as a relatively fixed judicial style.

Higher stringency strongly predicts higher incarceration rates for defendants in our sample. Panel A of Table 7 shows results from the first stage regression of an active sentence indicator on our judge stringency measure (Equation 6). For all results in column (D) of this table, we normalize judge stringency to standard deviation units within our direct exposure sample. The first

---

<sup>21</sup> We discuss the N.C. judicial process in more detail in Appendix B.

stage coefficient shows that a one standard deviation increase in our judicial stringency measure increases the probability that the defendant received an active sentence by 3.4 percentage points, or 24 percent of the mean. The  $F$ -statistic is 103, far exceeding typical rules of thumb of 10 or 16.38 (Staiger and Stock, 1997; Stock, Yogo, et al., 2005). Crucially, this relationship is not mechanical, as  $\mu_j$  is computed in a leave-out sample.<sup>22</sup>

We test the randomness of judicial assignment through balance tests, which suggest that the stringency of a defendant's judge is unrelated to their observable characteristics or the attributes of children in their household. In column (B) of Appendix Table A3, we regress  $\mu_j$  on defendant and case characteristics as well as attributes of the linked child (including offense class  $\times$  court  $\times$  disposition year fixed effects). We see no evidence of a meaningful correlation between our stringency measure and these characteristics. The joint  $F$ -statistic for all defendant/case characteristics is 1.1, and the joint  $F$ -statistic for all characteristics of the linked child is also 1.1.<sup>23</sup> This contrasts with the strong correlation between active sentences and defendant/child characteristics, which have joint  $F$ -statistics of 98 and 38 (column A of Appendix Table A3). Overall, these balance tests support our main identification assumption of quasi-randomness in judicial assignment.

## Results

We find that children's academic and behavioral outcomes are negatively impacted when a criminal defendant in their household is assigned to a more stringent judge. Column (D) of Table 7 shows estimates from the reduced form specification in Equation 5. The estimates in Panel B show that a one standard deviation increase in judge stringency reduces children's math and English scores by 0.6 and 1.2 percent of a standard deviation, respectively, although only the latter effect is statistically significant.<sup>24</sup> In Panel C of Table 7, we find that a one standard deviation increase in judge stringency raises the probability of suspension by 0.3 percentage points (2.4 percent of the mean), and it leads to marginally-significant increases in days of absence and fighting. These

---

<sup>22</sup> Appendix Figure A7 shows a fairly smooth and monotonically increasing relationship between  $\mu_j$  and active sentences in our sample.

<sup>23</sup> These estimates compare favorably, for example, with Mueller-Smith (2015), who finds  $F$ -statistics ranging from 0.9–2.3 for various defendant characteristics considered individually against judicial stringency. Panel B of Appendix Figure A7 shows one pairwise example of this balance, plotting a local linear regression of a key defendant characteristic—the presence of a prior offense—across the distribution of stringency.

<sup>24</sup> Panels C and D of Appendix Figure A7 display these results graphically, showing a negative association between judicial stringency and math and English scores.

reduced form estimates are generally smaller in magnitude than those in our event study (column (C)), but they are still economically meaningful given the wide variation in judge stringency.<sup>25</sup>

Column (E) of Table 7 presents estimates from the IV specification defined by Equations 3 and 6. The point estimates in this specification range from roughly 25–100 percent of the mean outcomes measured prior to disposition (column (A)). Dobbie et al. (2018) also find large negative effects on child outcomes in a similar IV specification, although the size of these estimates relative to the OLS coefficients suggests that the impacts of stringent judges may not operate exclusively through the incarceration decision.

Our OLS, event study, and judge stringency specifications rely on different sources of variation in incarceration that arises at different stages of the sentencing process. But each analysis yields a similar conclusion: direct exposure to incarceration in the household has adverse consequences on children’s academic and behavioral outcomes.

These findings add 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). Papers in this literature use both event study and judge designs, and vary in whether they find positive or negative impacts on children. The impacts of incarceration for children are likely to vary with the nature of the criminal activity and the child’s environment more broadly. The magnitudes of our event study and reduced-form stringency coefficients are relatively modest, and thus lie within the confidence intervals of the estimates in much of this work. For example, our math and English score effects are within the confidence intervals of Norris et al. (2021)’s estimates using a judge design in Ohio, and the authors find a positive but insignificant effects of incarceration on absences. (They do not have data on disciplinary incidents in school.) Further, our merge between students and defendants is based on addresses rather than birth records; many criminal defendants do not live with their children, and incarceration may be less disruptive to the home environment in these cases.

---

<sup>25</sup> We explore heterogeneity in the judge stringency effects on test scores in Panel B of Appendix Figure A9. Our results are largely consistent with our event study analysis; in particular, incarcerations of likely mothers are associated with large and negative impacts on children. Panel A of Appendix Table A4 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)).

## 5 Mechanisms: Indirect Classroom Exposure

Direct exposure to incarceration in the household has an economically meaningful impact on student achievement. However, given that a relatively small number of students are directly affected, our estimates are not large enough to explain the aggregate causal relationship between incarceration and test scores at the community level. To better explain the overall achievement-incarceration gradient, we next examine outcomes for students who are indirectly exposed to incarceration: the classmates of directly impacted children.

### 5.1 Empirical Specifications

To examine indirect impacts, we use the same empirical strategies as in Section 4, but we focus on our indirect exposure samples (see subsection 2.3 and columns (F)–(G) of Table 1). These samples include all 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 conduct each of our three empirical strategies—OLS, event study, and judge stringency—in the indirect exposure samples. Our regression specifications are similar to those in Section 4, but observations are defined at the classmate ( $k$ )  $\times$  directly-exposed child ( $i$ )  $\times$  calendar year ( $t$ ) level. For example, our event study regression for indirect impacts is:

$$y_{kit} = \delta \cdot \mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i + \eta_{ki} + \lambda_{og\tau(i)t} + \varepsilon_{kit}. \quad (7)$$

This specification is nearly identical to Equation 4, but observations are at the  $kit$  level and we include a fixed effect,  $\eta_{ki}$ , 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 just generalizations of Equations 3, 5, and 6 to the  $kit$  level.<sup>26</sup>

---

<sup>26</sup> Specifically, our OLS and reduced-form judge stringency specifications for indirect impacts are:

$$y_{kit} = \beta \cdot \text{Active Sentence}_i + \theta_{oct(i)} + \mathbf{X}_{kit}'\zeta + v_{kit} \quad (8)$$

$$y_{kit} = \gamma \cdot \mu_{j(i)} + \theta_{oct(i)} + \mathbf{X}_{kit}'\zeta + u_{kit}. \quad (9)$$

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 the directly-impacted child,  $i$  (both include characteristics of the defendant linked to child  $i$ ).

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

We present the same balance and robustness tests as in our direct exposure analysis. In Appendix Tables A2 and A4 (Panel B), we show that our event study and judge stringency estimates of indirect impacts are robust to alternate controls for offenses and defendant characteristics, and to alternate methods of defining stringency. Appendix Table A3 shows that judge stringency is not significantly related to characteristics of classmates in our indirect exposure sample (column (D)). These findings are consistent with the identification checks in our direct exposure analysis.

## 5.2 Results

We find consistent evidence that indirect exposure to household incarceration reduces children's academic achievement. Table 8 presents our main results on indirect impacts. This table is similar in structure to Table 7: 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 8, the OLS coefficients show a negative correlation between indirect exposure and test scores. The incarceration of a classmate's household member is associated with a reduction in math scores of 0.6 percent of a standard deviation and a reduction in English scores of 0.8 percent of a standard deviation. These estimates suggest that household incarceration may have meaningful spillover effects to classmates' academic performance, although they are subject to the usual concerns about omitted variable bias in cross-sectional comparisons.

Column (C) of Table 8 shows negative and precisely-estimated indirect impacts on test scores in our event study specification. The point estimates imply that the incarceration of a child's household member lowers the math and English scores of their classmates by 0.4 and 0.3 percent of a standard deviation, respectively. Panels A and B of Figure 3 show how these indirect impacts vary across years relative to the defendant's disposition using the same modified version of our event study as in Figure 2. While we observe mild declines in English scores (Panel B), the time

pattern of effects is particularly stark in math (Panel A); after completely flat pre-trends prior to the defendant's incarceration, we observe sizable declines in classmates' math scores that persist for three years.

Our judge stringency approach is not sufficiently 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 8 imply that a one standard deviation increase in the stringency of the judge linked to the directly-impacted child reduces their classmates' test scores by 0.1 percent of a standard deviation in both math and English.<sup>27</sup> Panels C and D of Figure 3 show the relationship between judicial stringency and these indirect impacts over time. We plot coefficients from a version of our reduced form specification that interacts  $\mu_{j(i)}$  with dummies for year relative to disposition (and includes outcomes in all years before and after disposition). Similar to the event study estimates in Panels A–B, these specifications show minimal pre-trends prior to disposition, and negative indirect impacts on test scores following the assignment of a stricter judge.

### 5.3 Channels of Indirect Impacts

Given our estimates in Section 4, which indicate that direct exposure to incarceration increases misbehavior, we posit that the disruptions generated by directly-impacted children may be a central mechanism for the indirect impacts on achievement. Following Lazear (2001)'s seminal model of classroom disruptions, there is now a large body of empirical evidence 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, Pashman, and Schlosser, 2012; Carrell et al., 2018; Billings and Hoekstra, 2019; Billings, Deming, and Ross, 2019).

The magnitudes of our estimates are in line with this literature. 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 indirect/direct

---

<sup>27</sup> The IV estimates in column (E) of Table 8 imply indirect impacts of a household incarceration of  $-0.02$  SDs in math and  $-0.05$  SDs in English. As in Table 7, the IV coefficients are significantly larger than the OLS coefficients, suggesting a possible violation of the IV assumptions.

ratio of incarceration impacts ranges from 10–27 percent, depending on the subject and empirical approach (event study vs. 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 ( $-0.025$ ), suggesting that the indirect effects of household incarceration on child behavior are not as substantial as the impacts of domestic violence.

Figure 4 provides evidence that math scores declined specifically for the classmates of the directly-impacted children whose misbehavior increased following household incarceration. For this, we take all children in our direct event study sample whose household member received an active sentence, and compute the change in *each* child's likelihood of suspension before versus after the disposition.<sup>28</sup> The *x*-axis of Figure 4 depicts this change in the likelihood of suspension, with each grey circle representing one directly-impacted child. The *y*-axis value of each circle represents the average change in math scores (from before to after the disposition) for the set of classmates linked to that child.<sup>29</sup> Red diamonds show average values in ventiles of the *x*-axis, and the dashed line shows the linear relationship between the *y*- and *x*-axis values. We find a negative relationship between the direct impact of incarceration on child misbehavior and its indirect impacts on math scores. On average, the negative effects on classmate math scores arise *only* in cases where the directly-impacted child's misbehavior increased.

Appendix Table A5 provides tests that formalize the results in Figure 4. To construct these tests, we first estimate our direct event study (equation 4) separately for *each* child  $i$  who was exposed to a household incarceration. We do this for each of our three behavioral outcomes (any suspension, suspension days, and fighting). This gives an individual-specific estimate of the direct effect on child behavior,  $\hat{\delta}_i$ , for each outcome. We then add the  $\hat{\delta}_i$  coefficients as a triple interaction term in our indirect event study (Equation 7), i.e.,  $\mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i \times \hat{\delta}_i$ . This interaction term tests whether classmate test scores changed differentially in cases where the directly-impacted child's misbehavior increased as the result of a household incarceration.

We find that test score impacts were worse—particularly in math—for the classmates of chil-

<sup>28</sup> To account for the fact that suspension rates increase with age, as well as any effects of the crime/arrest itself, we demean these individual-specific changes using the change in the likelihood of suspension for students in the same event study cell (offense  $\times$  academic cohort  $\times$  disposition year) whose household member did *not* receive an active sentence.

<sup>29</sup> Again, we demean changes in math scores using the average change among classmates linked to children in the same event study cell who were not exposed to a household incarceration.

dren whose misbehavior increased as a direct result of exposure to incarceration. Columns (A), (B), and (C) of Appendix Table A5 present results that include interactions with the change in the likelihood of suspension, number of suspension days, and likelihood of a fighting incidents. Panel A suggests that, across all three measures of misbehavior, the reduction in classmates' math scores was worse if the directly-impacted child began misbehaving more as a result of exposure. Panel B similarly shows negative estimates for English scores, although these effects are smaller and mostly statistically insignificant. The magnitudes imply that, for example, a 10 percentage point increase in the child's likelihood of suspension reduces their classmates' math scores by 0.1 percent of a standard deviation. These findings suggest that behavioral disruptions in the classroom are a mechanism by which incarceration indirectly impacts other students in the community.<sup>30</sup>

## 6 Discussion

This section discusses the implications of our direct and indirect results from Sections 4–5 for the community achievement-incarceration gradient in Figure 1. We illustrate the relative importance of the direct and indirect channels in Figure 5, with supporting calculations provided in Appendix Table A6. 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 7 and 8), as well the effects of adding demographic controls (Table 2).<sup>31</sup>

Figure 5 shows that the direct channel accounts for just 0.2 percent of the raw gradient between math scores and school incarceration exposure. 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 not enough direct exposures to account for much of the aggregate relationship.

---

<sup>30</sup> Appendix Figure A10 shows that the indirect impacts are more negative for classmates whose gender is the same as the directly-impacted child, suggesting the spillover effects may be stronger among students who are more likely to interact. Black students also appear to be more impacted by indirect exposure than white students.

<sup>31</sup> For this decomposition we assume constant effects at the mean level of incarceration exposure, and we ignore any dynamic effects. See Appendix Table A6 for details.

By contrast, the indirect channel can explain 9 percent of the raw achievement-incarceration gradient in math. The greater importance of the indirect channel is simply a function of the 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 A8). 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.<sup>32</sup> Thus while the indirect math score effects are small in magnitude (0.4 percent of a standard deviation), they aggregate to a larger total impact. Appendix Table A6 similarly shows that our indirect event study estimates for English scores can explain 6 percent of the raw gradient in English.

These results suggest that the direct impacts of incarceration on children—a key focus of prior research—are important for community-level achievement primarily because they indirectly impact these children’s classmates. Further, our analysis of mechanisms focused only on indirect exposures through children in the same school and grade, suggesting that other indirect causal channels may contribute to the raw gradient. Children interact across grade levels and outside of school, and may directly come in contact with incarcerated adults outside of their households. Figure 5 shows that demographic controls can explain 49 percent of the gradient (as in column C of Table 2), but a large portion remains unexplained. Our judge turnover results suggested that the remaining unexplained piece may largely be causal, and so we hope future research will shed further light on other channels of indirect impacts.

## 7 Conclusion

The central contribution of our paper is establishing that the negative impacts of incarceration have a broader reach into local communities than is conventionally assumed. We establish this result through causal estimation of community incarceration rates on the achievement of children who live in those communities, as well as through an analysis of the mechanisms that contribute to this result. We emphasize both direct impacts of incarceration on cohabitating children, as well as spillovers onto the classmates of those children.

---

<sup>32</sup> 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.

To explain broader community impacts, we find that the negative direct effects of incarceration on children include not only performance on standardized math and English tests, but also behavioral outcomes such as absences, suspensions, and fighting incidents. These behavioral effects negatively impact the academic performance of classmates, including those who do not experience an incarceration in their households. The spillover effects on children's achievement are persistent, lasting at least three years at both individual and aggregate levels.

When we calibrate the overall importance of the different mechanisms, we find that indirect exposure to household incarceration, though small in magnitude for any individual child, accounts for a large share of the overall gradient between achievement and community incarceration exposure. The strength of this indirect channel reflects the fact that children are so frequently exposed to incarceration, both at school and in their neighborhoods.

While our results are consistent with a large body of descriptive work that has documented how incarceration may negatively impact the incarcerated-individual and their family members, our work advances the literature with a research design that allows us to establish the causal effects of community-level incarceration rates. We also differ from prior work that uses judge designs in that we find negative effects of incarceration on children in the United States, which is globally unique in incarceration volume. Our findings point to the possible value of criminal justice reform that takes into account the broader consequences of mass incarceration. Incorporating the spillover effects of incarceration 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 J. Rockoff (2014). School segregation, educational attainment, and crime: Evidence from the end of busing in charlotte-mecklenburg. *The Quarterly Journal of Economics* 129(1), 435–476.
- 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., 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.
- 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.
- Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2021). On estimating multiple treatment effects with regression. arXiv:2106.05024.
- 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.
- Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Unpublished Working Paper.
- 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.

- 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*.
- Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica: journal of the Econometric Society*, 557–586.
- Stock, J. H., M. Yogo, et al. (2005). Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* 80(4.2), 1.
- 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

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
	Judge turnover sample			Direct exposure samples		Indirect exposure samples	
	All NC students	Not in court records	Linked to court records	Event study	Judge strin.	Event study	Judge strin.
<b>Panel A. Student characteristics</b>							
Male	0.51	0.51	0.51	0.50	0.50	0.51	0.51
Age	11.76	11.76	11.75	12.15	12.15	12.03	12.04
Age at disposition			11.42	12.89	12.85		
White	0.53	0.62	0.44	0.43	0.42	0.54	0.54
Black	0.27	0.20	0.33	0.39	0.40	0.26	0.26
Economically disadvantaged	0.51	0.41	0.62	0.72	0.72	0.49	0.49
<b>Panel B. Student outcomes</b>							
Math score	-0.04	0.11	-0.18	-0.32	-0.32	0.01	0.01
English score	0.00	0.15	-0.15	-0.28	-0.29	0.03	0.03
Days absent	7.31	6.72	7.97	8.79	8.80	7.10	7.10
Any suspension	0.08	0.06	0.11	0.14	0.14	0.08	0.08
Number of suspension days	0.76	0.56	0.98	1.56	1.55	0.86	0.86
Fighting incident	0.024	0.018	0.032	0.045	0.045	0.025	0.025
<b>Panel C. Defendant/offense characteristics</b>							
Male			0.64	0.73	0.70		
Age at disposition			32.88	32.72	32.75		
White			0.49	0.48	0.47		
Black			0.36	0.42	0.43		
Felony offense			0.18	0.34	0.32		
Misdemeanor offense			0.23	0.38	0.35		
Traffic offense			0.59	0.27	0.33		
Guilty verdict			0.46	1.00	0.79		
Active sentence			0.15	0.29	0.14		
# students	2,191,669	1,186,040	891,764	128,829	118,416	1,470,600	1,454,577
# Math scores	3,903,718	1,763,568	1,834,605	699,084	643,595	6,951,743	6,905,245
# defendants			494,285	86,854	79,765		

*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 our event study sample for the direct impacts of household incarceration, which includes students whose linked defendant was convicted of their most serious offense. Column (E) 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. Columns (F)–(G) include students who were in the same school and grade as the children in columns (D)–(E) in the year of the defendant’s disposition; these are our samples for examining the indirect impacts of 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 be mean zero and standard deviation one in the full population of test takers in each year. In Panel B, student outcomes are measured from 2010–2014 in columns (A)–(C), and from 2006–2017 in columns (D)–(G). The defendant/offense characteristics in Panel C correspond to the most serious offense linked to each child, as described in the text. See Appendix C.1 for details on variable definitions.

TABLE 2: OLS effects 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
<b>Panel A. School exposure: Linear effect of 1 log point in HH incarcerations/student</b>					
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
<b>Panel B. Neighborhood exposure: Linear effect of 1 log point in incarcerations/resident</b>					
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
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,  $\gamma_t$ , 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: County stringency balance tests and effects on active sentencing

Dependent variable	2010 mean (levels)	(A)		(B)		(C)	
		Coefficient on $\Delta$ county stringency (SD units)					
		Actual caseloads		Fixed caseloads			
		Coef	(SE)	Coef	(SE)		
<b>Panel A. Active sentences</b>							
$\Delta \log \#$ active sentences	622	0.208***	(0.067)	0.166**	(0.064)		
<i>F</i> statistic		9.7		6.6			
<b>Panel B. Criminal case totals</b>							
$\Delta \log \#$ total cases	6,802	0.006	(0.013)	0.023*	(0.013)		
$\Delta \log \#$ felony cases	904	0.008	(0.030)	0.020	(0.037)		
$\Delta \log \#$ misdemeanor cases	3,279	-0.004	(0.018)	0.001	(0.015)		
$\Delta \log \#$ traffic cases	2,379	0.015	(0.021)	0.049	(0.031)		
$\Delta \log \#$ clerk to decide cases	240	0.041	(0.078)	0.040	(0.108)		
<i>F</i> statistic: All coefficients zero		0.3		1.5			
<b>Panel C. Defendant characteristics</b>							
$\Delta$ proportion male	0.704	-0.001	(0.002)	0.004	(0.003)		
$\Delta$ mean age	32.420	-0.051	(0.047)	-0.003	(0.068)		
$\Delta$ proportion white	0.476	0.003	(0.006)	0.004	(0.007)		
$\Delta$ proportion Black	0.433	-0.003	(0.007)	-0.003	(0.007)		
$\Delta$ proportion w/ multiple offenses	0.432	-0.007*	(0.004)	-0.005	(0.004)		
$\Delta$ proportion w/ prior offense	0.505	-0.001	(0.005)	-0.001	(0.005)		
$\Delta$ 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			
<b>Panel D. Student characteristics</b>							
$\Delta$ proportion male	0.507	-0.000	(0.001)	0.000	(0.001)		
$\Delta$ mean age	11.764	0.005	(0.008)	-0.001	(0.011)		
$\Delta$ proportion white	0.541	-0.002	(0.001)	-0.002*	(0.001)		
$\Delta$ proportion Black	0.270	0.003**	(0.001)	0.003	(0.002)		
$\Delta$ proportion economically disadvantaged	0.492	0.001	(0.002)	0.002	(0.002)		
<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,  $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$ .  $\Delta S_{ct}$  is the change in county stringency using both actual caseloads (column B) and fixed caseloads (column C), and  $\gamma_t$  are year fixed effects. Our outcomes,  $\Delta Y_{ct}$ , measure changes in four sets of variables. In Panel A, the outcome variable is changes in log total active sentences at the county level. Panel B investigates changes in county-level criminal case totals. 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). 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 4: Effects of county stringency on student test scores

Dependent variable	2010 mean (levels)	(A)	(B)	(C)	(D)	(E)
		Coefficient on $\Delta$ county stringency (SD units)				
		All students		Students not linked to court records		
Panel A. Academic performance		Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads	
$\Delta$ mean Math score	-0.031	-0.024*** (0.008)	-0.036** (0.018)	-0.027*** (0.007)	-0.045** (0.017)	
$\Delta$ mean English score	0.001	-0.014** (0.007)	-0.020* (0.010)	-0.015** (0.006)	-0.026*** (0.009)	
N (# county/years)	100	400	400	400	400	
# Math scores	778,942	3,903,718	3,903,718	1,763,568	1,763,568	
Panel B. Individual year-to-year change in test scores						
Mean individual $\Delta$ Math score	0.015	-0.025 (0.015)	-0.025* (0.015)	-0.028* (0.016)	-0.032** (0.016)	
Mean individual $\Delta$ Reading score	0.008	-0.013 (0.010)	-0.018** (0.009)	-0.013 (0.010)	-0.021** (0.008)	
N (# county/years)	100	400	400	400	400	
# Math scores	533,502	2,734,802	2,734,802	1,217,082	1,217,082	

*Notes:* This table shows the impacts of changes in county stringency from 2010–2014, measured using a judge turnover strategy, on changes student test scores. We follow Equation 2 and estimate the regression:  $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$ .  $\gamma_t$  are year fixed effects.  $\Delta S_{ct}$  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:  $\Delta S_{ct} = \sum_j (\omega_{jct} - \omega_{jct-1}) \mu_{jct}^{-\{t-1,t\}}$ . Columns (B) and (D) show results using the actual caseload weights  $\omega$  which judges face, and columns (C) and (E) show results using a sample which keeps caseloads fixed (in this specification, variation only results from judges entering and leaving the sample). Columns (D)–(E) restrict to students who never match to a defendant in the court records (column B of Table 1). Across all specifications, we use a jackknife leave-out estimate of judicial stringency  $\mu_{jct}^{-\{t-1,t\}}$ . We scale  $\Delta S_{ct}$  so that one unit represents one standard deviation of the distribution of county stringency. In Panel A, outcome variables are Math and English scores in standard deviation units, defined as in Table 1. In Panel B, outcomes variables measure year-to-year changes in each student's own Math and English; the sample thus excludes any students who did not take Math or English exams in consecutive years. Parentheses contain standard errors clustered at the county level.

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

TABLE 5: Leading and lagged effects of a change in county stringency between years  $t - 1$  and  $t$ 

Timing of dependent variable	(A)		(B)		(C)		(D)		(E)		(F)	
	Dep. variable: Log # active sentences		Dep. variable: Mean Math score		Dep. variable: Mean English score							
	Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads						
$\Delta Y_{t-3,t-1}$	0.031 (0.039)	-0.012 (0.070)	0.017 (0.013)	0.028* (0.016)	0.005 (0.010)	0.012 (0.011)						
$\Delta Y_{t-2,t-1}$	0.055 (0.056)	0.021 (0.073)	0.009 (0.009)	0.016 (0.010)	0.005 (0.008)	0.001 (0.008)						
$\Delta Y_{t,t-1}$	0.208*** (0.067)	0.166** (0.064)	-0.024*** (0.008)	-0.036** (0.018)	-0.014** (0.007)	-0.020* (0.010)						
$\Delta Y_{t+1,t-1}$	0.261*** (0.072)	0.204** (0.082)	-0.051*** (0.015)	-0.061*** (0.022)	-0.029*** (0.009)	-0.033** (0.015)						
$\Delta Y_{t+2,t-1}$	0.264*** (0.074)	0.208*** (0.072)	-0.040* (0.022)	-0.055** (0.021)	-0.027** (0.012)	-0.037*** (0.011)						
N (# county/years)	400	400	400	400	400	400						
# test scores ( $\Delta Y_{t,t-1}$ )			3,903,718	3,903,718	4,016,529	4,016,529						
<i>F</i> -statistics:												
All prior coeffs zero	0.6	0.1	0.9	1.7	0.2	0.7						
All post coeffs zero	4.5	2.9	4.2	2.7	3.2	3.9						

*Notes:* This table reports a modified version of Equation 2 from Table 4 to examine leading and lagged effects of a change in county stringency. We run the specification:  $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$ , but change the timing of the dependent variable relative to the year of the county stringency shock. The top two rows consider differences in outcomes between  $t - 3$  and  $t - 1$  as well as  $t - 2$  and  $t - 1$  to evaluate trends prior to time  $t$ . The middle row considers the difference in outcomes between  $t$  and  $t - 1$ , which replicates results from Table 4. The bottom two rows examine differences in outcomes between  $t + 1$  and  $t - 1$  as well as  $t + 2$  and  $t - 1$  to evaluate the longer-run impacts of a change at time  $t$ . In all cases,  $\Delta S_{ct}$  is defined as the change in county stringency between year  $t$  and  $t - 1$ . Columns (A)–(B) measure outcomes of changes in the log county-level active sentences across these time periods, broken out into the actual (column (A)) and fixed caseload measure (column (B)). Columns (C)–(F) similarly examine changes in student math and English scores in standard deviation units, defined as in Table 1. Parentheses contain standard errors clustered at the county level. The bottom of the table reports *F*-statistics from joint significance tests of all the prior ( $t - 3$  and  $t - 2$ ) and post ( $t$ ,  $t + 1$ , and  $t + 2$ ) period coefficients.

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

TABLE 6: Sources of variation in county stringency

Dependent variable	(A)	(B)	(C)	(D)	(E)
	Actual caseloads	Fixed caseloads	Judicial elections	Mid-term leaves	County rotations
<b>Panel A. Variance decomposition for change in county stringency (<math>\Delta S_{st}</math>)</b>					
SD of change in county stringency ( $\Delta S_{ct}$ )	0.0180	0.0140	0.0077	0.0064	0.0032
Variance of $\Delta S_{ct}$	32.5E-5	19.6E-5	5.9E-5	4.1E-5	1.0E-5
% of total variance in $\Delta S_{ct}$	100%	60%	18%	13%	3%
<b>Panel B. Main effects (scaled to 1 SD of county stringency in each measure)</b>					
$\Delta$ log # active sentences	0.208*** (0.067)	0.166** (0.064)	0.102 (0.071)	0.270*** (0.102)	0.207 (0.172)
$\Delta$ mean Math score	-0.024*** (0.008)	-0.036** (0.018)	-0.035* (0.018)	-0.100** (0.044)	0.042 (0.052)
$\Delta$ mean English score	-0.014** (0.007)	-0.020* (0.010)	-0.027** (0.013)	-0.046** (0.023)	0.033 (0.047)
N (# county/years)	400	400	400	400	400
# Math scores	3,903,718	3,903,718	3,903,718	3,903,718	3,903,718
<b>Panel C. Balance tests (F statistics)</b>					
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,  $\Delta S_{ct}$ , 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  $\Delta S_{ct}$  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. Lastly, 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,  $\Delta S_{ct}$  (first row). This panel also reports the percent of the total variance in  $\Delta S_{ct}$  explained by each component, which is equal to the variance of  $\Delta S_{ct}$  in each column divided by the variance of  $\Delta S_{ct}$  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 Table 4, but we use  $\Delta S_{ct}$  computed from the source of variation highlighted in each column. We normalize each  $\Delta S_{ct}$  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 3 and 5, but again use  $\Delta S_{ct}$  computed from the source of variation for each column.

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

TABLE 7: Direct impacts of household incarceration on student outcomes

Dependent variable	Mean	(A)	(B)	(C)	(D)	(E)
		OLS coef. on active sentence	Event study estimate	Judge stringency estimate	Reduced form (SD units)	IV for active sentence
<b>Panel A. First stage</b>						
Active sentence	0.143				0.034*** (0.003)	
<b>Panel B. Academic performance</b>						
Math score	-0.282	-0.034*** (0.010)	-0.015*** (0.005)	-0.006 (0.006)	-0.164 (0.169)	
English score	-0.265	-0.014 (0.011)	-0.012** (0.005)	-0.012** (0.006)	-0.348** (0.164)	
<b>Panel C. Attendance and misbehavior</b>						
Days absent	8.365	0.412*** (0.094)	0.456*** (0.059)	0.079* (0.045)	2.372* (1.396)	
Any suspension	0.123	0.009*** (0.003)	0.013*** (0.002)	0.003** (0.001)	0.099** (0.043)	
Number of suspension days	1.410	0.153*** (0.055)	0.252*** (0.036)	0.021 (0.022)	0.642 (0.666)	
Fighting incident	0.040	0.003** (0.002)	0.002* (0.001)	0.001* (0.001)	0.042* (0.024)	
Judge stringency sample	Y	Y	Y	Y	Y	Y
Event study sample			Y			
Years, $t$ , relative to case disposition, $\tau(i)$	$t < \tau(i)$	$t \geq \tau(i)$	All $t$	$t \geq \tau(i)$	$t \geq \tau(i)$	
N (Math scores)	339,549	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		
Child balance test				1.1		

*Notes:* This table presents estimates of the direct impacts of a household incarceration using both our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our direct judge stringency sample (column (E) of Table 1). Column (C) includes students in our direct event study sample (column D of Table 1). 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,  $t$ , before the disposition year of the defendant's case,  $\tau(i)$ . Column (B) shows OLS estimates of  $\beta$  from Equation 3 using outcomes measured in or after the disposition year ( $t \geq \tau(i)$ ). Column (C) displays estimates of  $\delta$  from the event study specification 4 using all years  $t$ . Column (D) shows reduced-form judge stringency estimates of  $\gamma$  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 IV estimates of  $\beta$  from the IV specification 3 and 6 for 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 (columns C) and judge (columns (B), (D), and (E)) levels. The bottom of the table reports report *F*-statistics from the first stage and joint significance tests for defendant and student characteristics for the specification in column (D); see Appendix Table A3 for details on the balance tests.

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

TABLE 8: Indirect impacts of household incarceration on classmates

Dependent variable	Mean	(A)	(B)	(C)	(D)	(E)
		OLS coef. on active sentence	Event study estimate	Judge stringency estimate	Reduced form (SD units)	IV for active sentence
<b>Panel A. First stage</b>						
Active sentence	0.136				0.031*** (0.002)	
<b>Panel B. Academic performance</b>						
Math score	-0.001	-0.006* (0.003)	-0.004** (0.002)	-0.001 (0.002)	-0.017 (0.062)	
English score	0.002	-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, $\tau(i)$		$t < \tau(i)$	$t \geq \tau(i)$	All $t$	$t \geq \tau(i)$	$t \geq \tau(i)$
N (Math scores)	62,008,644	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	
F 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 using both our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our indirect judge stringency sample (column G of Table 1). Column (C) includes students in our indirect event study sample (column F of Table 1). 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-impact child in their school/grade.

Column (A) displays means of each outcome variable measured calendar years,  $t$ , before the disposition year of the defendant’s case,  $\tau(i)$ . Column (B) shows OLS estimates of  $\beta$  from equation 8 using outcomes measured in or after the disposition year ( $t \geq \tau(i)$ ). Column (C) displays estimates of  $\delta$  from the event study specification 7 using all years  $t$ . Column (D) shows reduced-form judge stringency estimates of  $\gamma$  from equation 9 for years  $t \geq \tau(i)$ , with estimates normalized to represent a one standard deviation increase in judge stringency. Column (E) shows IV estimates of  $\beta$  from equation 8 for years  $t \geq \tau(i)$ , where we use judge stringency as an instrument for Active Sentence $_i$ . 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 cluster at the school level in all columns (B)–(E). The bottom of the table reports report  $F$ -statistics from the first stage and joint significance tests for defendant and student characteristics for the specification in column (D); see Appendix Table A3 for details on the balance tests.

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

## Figures

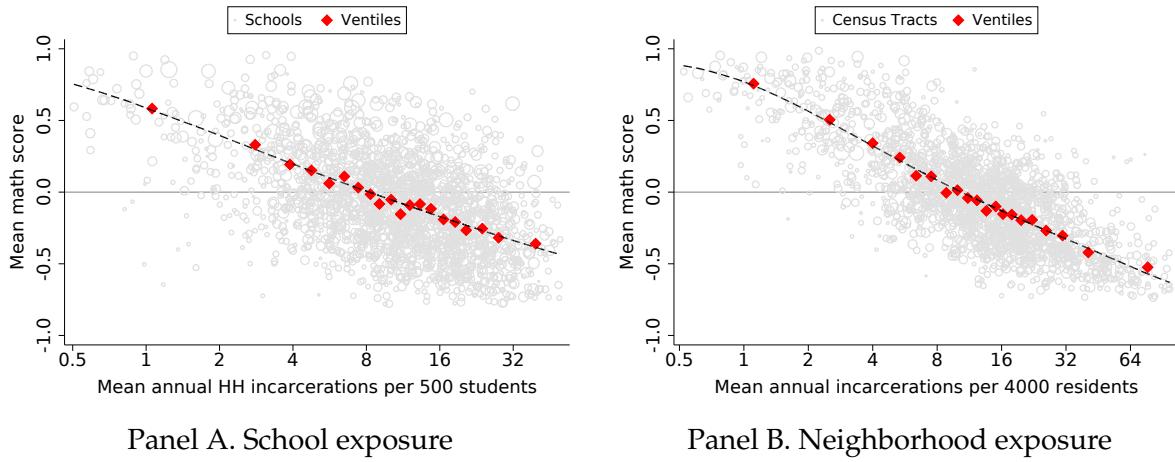


FIGURE 1: Math scores by school and neighborhood exposure to incarceration events (log scale)

*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, grey 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, grey 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 be mean zero and standard deviation 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 99<sup>th</sup> and 1<sup>st</sup> percentiles, and use log scales on the x axis excluding observations below 0.5 mean annual incarcerations per 500 students/4,000 residents.

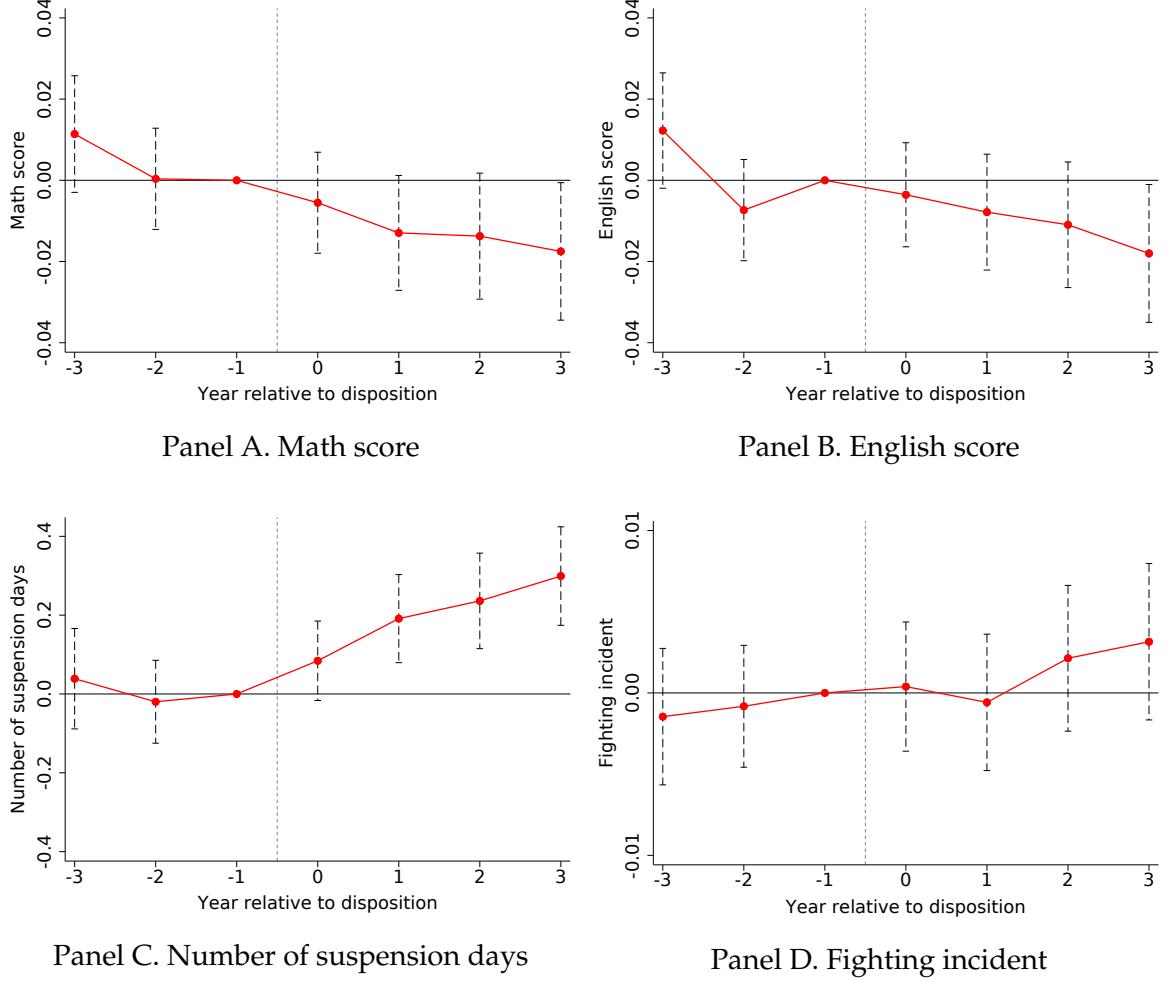


FIGURE 2: Direct impacts of household incarceration — Event study

*Notes:* This figure shows event studies of the direct impacts of household incarceration on students' academic and behavioral outcomes. We use our direct exposure event study sample (column D in Table 1) which focuses on students who have a convicted household member, and use variation in whether this family member received an active sentence conditional on granular controls. We follow equation 4:  $y_{it} = \sum_{l=-8}^9 \delta_l \mathbb{1}\{t - \tau(i) = l\} \times \text{Active Sentence}_{k(i)} + \eta_i + \lambda_{og\tau(i)t} + \varepsilon_{it}$ . These specifications show outcomes  $y_{it}$  for student  $i$  in calendar year  $t$  who has a household member  $k(i)$  that receives a disposition in year  $\tau(i)$ .  $\text{Active Sentence}_{k(i)}$  is an indicator equal to one if defendant  $k(i)$  received an active sentence.  $\eta_i$  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  $\delta_l$  coefficients, where the  $x$ -axis denotes years  $l$  relative to the defendant's disposition. We omit the  $l = -1$  interaction term, so  $\delta_l$  represents the difference in outcomes between students with and without a household incarceration in year  $l$  relative to the same difference measured one year before the disposition. We plot only outcomes measured from three years prior to three years after the disposition year.

The outcome variables,  $y_{it}$ , are Math (Panel A) and English (Panel B) scores in standard deviation units, total number of days with a suspension (Panel C), and an indicator for any fighting incident during the year (Panel D). Dashed lines are 95 percent confidence intervals using standard errors clustered at the defendant level.

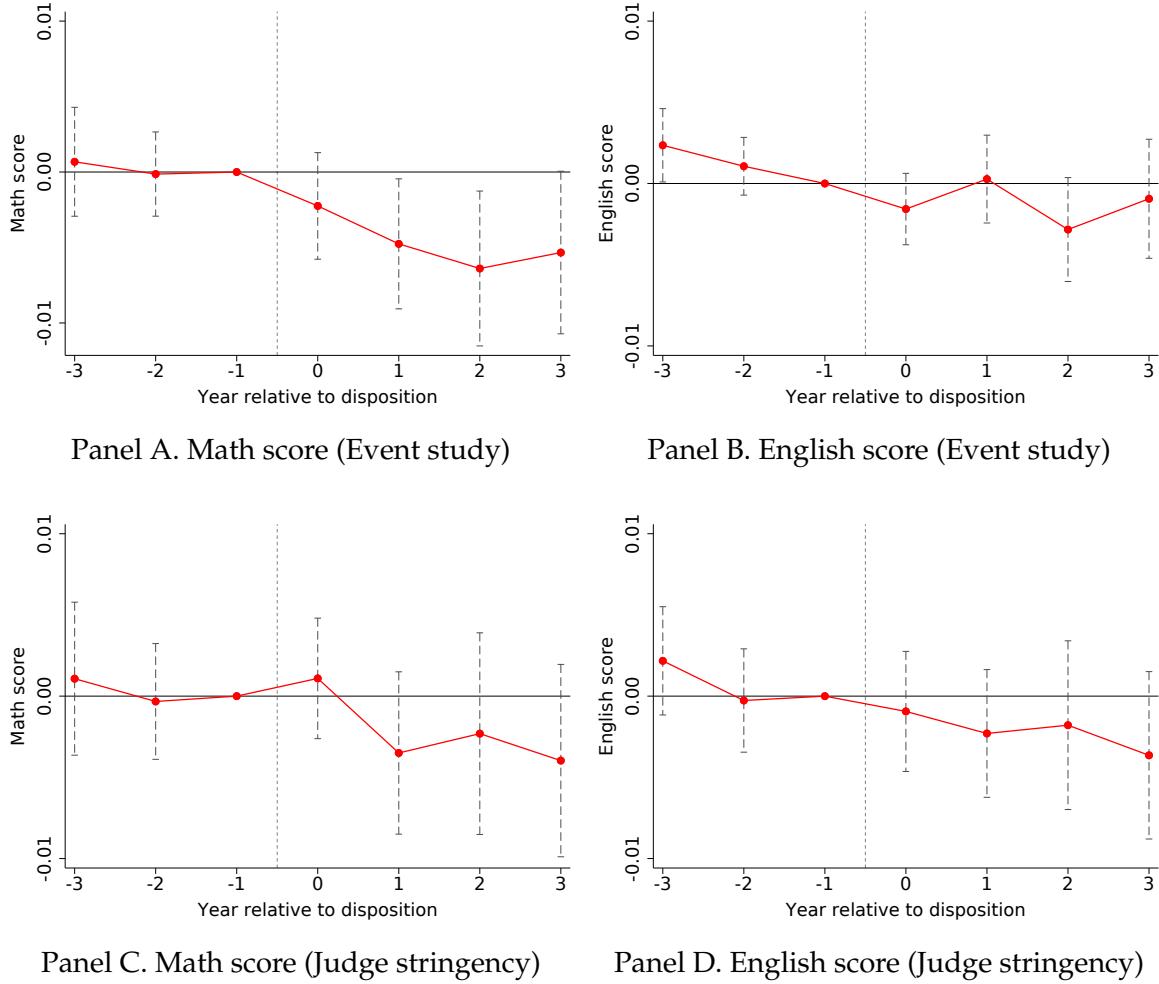


FIGURE 3: Indirect impacts of household incarceration — Event study and timing of judge stringency effects

Notes: This figure shows event study and judge stringency estimates of the indirect impacts of household incarceration on students' academic outcomes.

Panels A–B present event study estimates using our indirect exposure sample from column (F) in Table 1. We follow the dynamic version of equation 7:  $y_{ikt} = \sum_{l=-8}^9 \delta_l \cdot \mathbb{1}\{t - \tau(k) = l\} \times \text{Active Sentence}_{d(k)} + \eta_i + \lambda_{og\tau(k)t} + \varepsilon_{ikt}$ . These specifications show outcomes  $y_{ikt}$  for student  $i$  in calendar year  $t$  whose classmate  $k$  has a household member  $d(k)$  that receives a disposition in year  $\tau(k)$ .  $\text{Active Sentence}_{d(k)}$  is an indicator equal to one if defendant  $d(k)$  received an active sentence.  $\eta_i$  is an individual fixed effect.  $\lambda_{og\tau(k)t}$  is an offense class  $\times$  academic cohort  $\times$  disposition year  $\times$  calendar year fixed effect. The graphs plot the  $\delta_l$  coefficients, where the  $x$ -axis denotes years  $l$  relative to the defendant's disposition. We omit the  $l = -1$  interaction term, so  $\delta_l$  represents the difference in outcomes between students with and without a household incarceration in year  $l$  relative to the same difference measured one year before the disposition.

Panels C–D present judge stringency estimates using our indirect exposure sample from column (G) in Table 1. We estimate our reduced-form judge stringency specification from Equation 9, and interact the stringency measure with dummies for years  $l$  relative to the defendant's disposition. The graphs plot the  $\gamma_l$  coefficients with years since disposition,  $l$ , on the  $x$ -axis. We omit the  $l = -1$  interaction term, so  $\gamma_l$  represents the indirect impacts of receiving a more stringent judge in year  $l$  relative to the same effect measured one year before the disposition. We normalize estimates to represent a one standard deviation increase in judge stringency.

The outcome variables,  $y_{ikt}$ , are Math (Panels A and C) and English (Panels B and D) scores in standard deviation units. We plot only outcomes measured from three years prior to three years after the disposition year. Dashed lines in all panels are 95 percent confidence intervals using standard errors clustered at the school level.

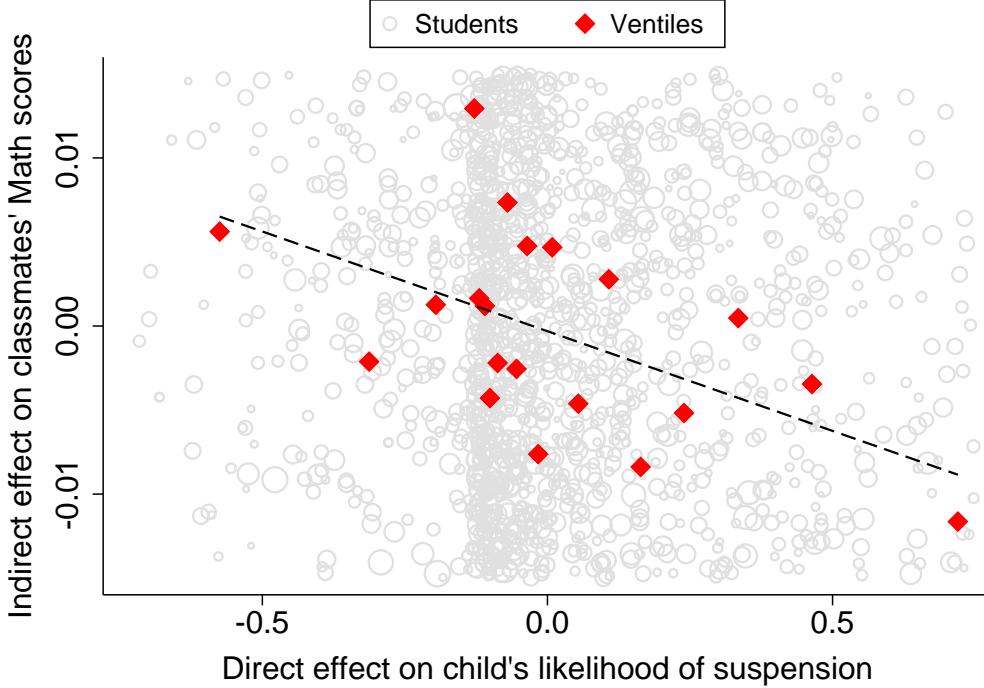


FIGURE 4: Classroom disruption channel of indirect academic impacts

*Notes:* This figure plots the relationship between the direct impacts of household incarceration on the likelihood of a suspension ( $x$ -axis) and the indirect impacts on the math scores of each student's classmates ( $y$ -axis).

On the  $x$ -axis, each grey circle represents a student in our direct exposure event study sample (column D of Table 1) whose household member received an active sentence. The  $x$ -axis value is the change in the directly-impacted student's likelihood of receiving a suspension from before to after the defendant's disposition (i.e., the difference in the proportion of years with a suspension). We demean this value using the average change in the likelihood of a suspension for students in the same event study cell (offense class  $\times$  academic cohort  $\times$  disposition year) as the directly-impacted child, but whose household member was convicted *without* an active sentence. Thus if we average the demeaned changes ( $x$ -axis values) across directly-impacted students, we replicate our main event study estimate for "any suspension" in Table 7 (0.013).

On the  $y$ -axis, each grey circle represents the *set* of students in our indirect exposure event study sample (column F of Table 1) who were in the same school/grade as the directly-impacted child in the disposition year (i.e., the directly-impacted student's "classmates"). The  $y$ -axis value is the change in these classmates' math scores from before to after the defendant's disposition. We demean this value using the average change in math scores for students in the same event study cell (offense class  $\times$  academic cohort  $\times$  disposition year) as the directly-impacted child's classmates, but who were instead classmates of students whose household member was convicted *without* an active sentence. Thus if we average the demeaned changes ( $y$ -axis values) across the directly-impacted students' classmates, we replicate our main event study estimate for the indirect math score effect in Table 8 (-0.004).

The dashed line shows the OLS relationship between changes in classmates' math scores ( $y$ -axis) and changes in directly-impacted students' suspension rate ( $x$ -axis). Red diamonds plots means of each variable in ventiles of the  $x$ -axis values.

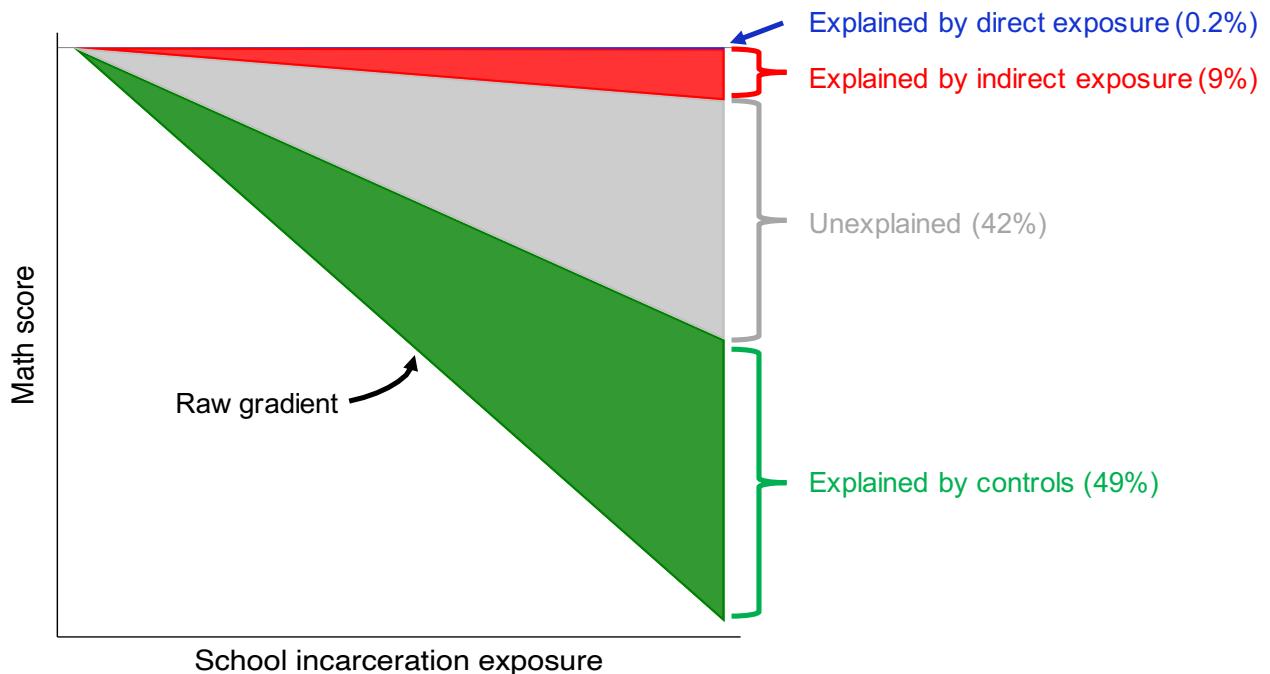


FIGURE 5: Decomposition of the gradient between math scores and school incarceration exposure

*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 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 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 A6 for details on these calculations. The grey region shows the remaining unexplained component (42 percent of the raw gradient).

## A Appendix Tables and Figures

TABLE A1: Heterogeneity in county stringency effects by exposure to criminal cases

	(A)	(B)	(C)	(D)	(E)					
	Coefficient on $\Delta$ county stringency (SD units)									
<b>Panel A. Heterogeneity by school exposure to criminal cases</b>										
Quartiles of criminal cases per student in 2010										
Dependent variable	All schools	Bottom Qtile	Q2	Q3	Top Qtile					
$\Delta \log \text{HH incarcerations per student}$	0.126*** (0.038)	0.121 (0.141)	0.139* (0.079)	0.075 (0.051)	0.296** (0.109)					
$\Delta \text{HH incarcerations per 500 students}$	1.031** (0.485)	0.562 (1.317)	1.481 (1.124)	0.187 (0.456)	4.657*** (1.475)					
$\Delta \text{mean Math score}$	-0.028*** (0.008)	-0.025 (0.048)	-0.015 (0.009)	-0.027** (0.009)	-0.059** (0.024)					
$\Delta \text{mean English score}$	-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					
<b>Panel B. Heterogeneity by neighborhood exposure to criminal cases</b>										
Quartiles of criminal cases per resident in 2010										
Dependent variable	All tracts	Bottom Qtile	Q2	Q3	Top Qtile					
$\Delta \log \text{incarcerations per resident}$	0.142*** (0.031)	0.172** (0.070)	0.132*** (0.033)	0.138*** (0.036)	0.095** (0.040)					
$\Delta \text{incarcerations per 4000 residents}$	2.332*** (0.546)	1.254*** (0.279)	1.897*** (0.405)	2.321*** (0.689)	3.622** (1.742)					
$\Delta \text{mean Math score}$	-0.026*** (0.008)	-0.019** (0.009)	-0.026** (0.011)	-0.024** (0.011)	-0.028*** (0.009)					
$\Delta \text{mean English score}$	-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  $\beta$  from Equation 2, but outcomes are defined at the school (Panel A) or Census tract (Panel B) level rather than 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 A2: Event study robustness tests

Dependent variable	(A) Benchmark	(B) Offense class at arrest	(C) 4-digit offense code	(D) Controls for prior points
<b>Panel A. Direct effects of household incarceration</b>				
Math score	-0.015*** (0.005)	-0.016*** (0.005)	-0.013** (0.006)	-0.014** (0.006)
English score	-0.012** (0.005)	-0.015*** (0.005)	-0.011* (0.006)	-0.009* (0.005)
Days absent	0.456*** (0.059)	0.469*** (0.057)	0.322*** (0.065)	0.435*** (0.060)
Any suspension	0.013*** (0.002)	0.015*** (0.002)	0.009*** (0.002)	0.011*** (0.002)
Number of suspension days	0.252*** (0.036)	0.260*** (0.036)	0.201*** (0.040)	0.193*** (0.038)
Fighting incident	0.002* (0.001)	0.003** (0.001)	0.001 (0.001)	0.001 (0.001)
N (Math scores)	696,289	696,398	655,070	696,289
<b>Panel B. Indirect effects of household incarceration</b>				
Math score	-0.004** (0.002)	-0.006*** (0.002)	-0.004* (0.002)	-0.004* (0.002)
English score	-0.003** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)	-0.002* (0.001)
N (Math scores)	103,307,344	103,307,354	103,298,159	103,307,344

*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  $\delta$  from the direct event study regression 4. Panel B presents estimates of  $\delta$  from the indirect event study regression 7. Column (A) presents estimates from our benchmark specification, which replicates the results in column (C) of Tables 7 and 8. Column (B) defines offenses,  $o$ , by the offense class at arrest rather than the offense class at conviction (as in our benchmark). Column (C) defines offenses,  $o$ , by the 4-digit offense code at conviction rather than the offense class at conviction. Column (D) is identical to column (A), 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 A3: 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)
<b>Panel A. Defendant/offense characteristics</b>				
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
<b>Panel B. Student characteristics</b>				
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
N	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 E of Table 1). Regressions in columns (C)–(D) are at the directly-impacted student × classmate level using all classmates in our indirect judge stringency sample (column G of Table 1). All regressions include court × year × 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.

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

TABLE A4: 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
<b>Panel A. Direct effects of household incarceration</b>					
Active sentence	0.034*** (0.003)	0.033*** (0.003)	0.022*** (0.002)	0.028*** (0.004)	0.053*** (0.009)
Math score	-0.006 (0.006)	-0.009 (0.005)	-0.006* (0.003)	-0.002 (0.007)	-0.021* (0.012)
English score	-0.012** (0.006)	-0.012** (0.005)	-0.009*** (0.003)	-0.012 (0.008)	-0.023* (0.012)
Days absent	0.079* (0.045)	-0.026 (0.045)	0.002 (0.028)	0.032 (0.055)	0.306** (0.120)
Any suspension	0.003** (0.001)	0.003* (0.001)	0.002** (0.001)	0.004** (0.002)	0.015*** (0.003)
Number of suspension days	0.021 (0.022)	0.012 (0.021)	0.013 (0.013)	0.020 (0.026)	0.059 (0.054)
Fighting incident	0.001* (0.001)	0.001 (0.001)	0.001* (0.000)	0.001 (0.001)	0.001 (0.002)
N (Math scores)	295,905	285,192	303,712	294,331	84,431
<b>Panel B. Indirect effects of household incarceration</b>					
Active sentence	0.031*** (0.002)	0.030*** (0.002)	0.020*** (0.001)	0.026*** (0.002)	0.047*** (0.008)
Math score	-0.001 (0.002)	-0.002 (0.002)	-0.002* (0.001)	-0.002 (0.002)	-0.006 (0.004)
English score	-0.001 (0.002)	-0.003* (0.002)	-0.002* (0.001)	-0.004* (0.002)	-0.005 (0.003)
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 household incarceration. Panel A presents reduced form estimates of  $\gamma$  from the direct judge stringency regression 5. Panel B presents reduced form estimates of  $\gamma$  from the indirect judge stringency regression 9. Column (A) presents estimates from our benchmark specification, which replicates the results in column (D) of Tables 7 and 8. For column (B), we compute stringency at the judge  $\times$  year level rather than 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  $\times$  year  $\times$  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  $\times$  year level. Column (D) is identical to column (A), except that in the fixed effect term,  $\theta_{oct}$ , we define offenses,  $o$ , using 4-digit offense code rather than 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 A5: Heterogeneity in indirect event study estimates by direct effects on child behavior

Independent variables	Behavioral outcome for directly-impacted child		
	Any suspension	Number of suspension days	Fighting incident
<b>Panel A. Dependent variable: Math scores of indirectly-impacted children</b>			
Active sentence $\times$ Post disposition	-0.0040** (0.0019)	-0.0039** (0.0019)	-0.0041** (0.0019)
Active $\times$ Post $\times$ Direct effect on any suspension	-0.0112*** (0.0042)		
Active $\times$ Post $\times$ Direct effect on # suspension days		-0.0005*** (0.0002)	
Active $\times$ Post $\times$ Direct effect on fighting incident			-0.0112* (0.0065)
N (Math scores)	102,146,692	102,146,692	102,146,692
<b>Panel B. Dependent variable: English scores of indirectly-impacted children</b>			
Active sentence $\times$ Post disposition	-0.0025** (0.0013)	-0.0026** (0.0013)	-0.0026** (0.0013)
Active $\times$ Post $\times$ Direct effect on any suspension	-0.0052* (0.0029)		
Active $\times$ Post $\times$ Direct effect on # suspension days		-0.0001 (0.0001)	
Active $\times$ Post $\times$ Direct effect on fighting incident			-0.0070 (0.0045)
N (English scores)	107,314,119	107,314,119	107,314,119

*Notes:* This table examines heterogeneity in the *indirect* effects of household incarceration on classmates' test scores. We consider heterogeneity based on the *direct* effects of household incarceration on child behavior.

For this, we first estimate the direct effects of household incarceration on the behavior of *each* directly-impacted child. We use our direct event study specification 4, but we estimate this regression separately for each child in our direct sample whose linked defendant received an active sentence. We estimate these regressions for three behavior outcomes: any suspension, number of suspension days, and any fighting incident. This gives individual-specific estimates of the effects of household incarceration on child behavior, which we denote here by  $\hat{\delta}_i$ . If we average these  $\hat{\delta}_i$  estimates, we approximately reproduce our main direct event study estimates in column (C) of Table 7.

We then add these individual-specific direct effects as an interaction term in our indirect event study specification. This specification is identical to Equation 7, except we include the triple interaction term  $\mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i \times \hat{\delta}_i$ . This table displays the coefficient on both the double interaction,  $\mathbb{1}\{t \geq \tau(i)\} \times \text{Active Sentence}_i$ , and the triple interaction term. In Panel A, the dependent variable is the classmate's Math score. In Panel B, the dependent variable is the classmate's English score. Columns (A)–(C) present results from separate regressions using the three different behavioral outcomes to define  $\hat{\delta}_i$ . Parentheses contain standard errors clustered at the school level.

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

TABLE A6: Decomposition of the gradient between test scores and school incarceration exposure

	(A)	(B)	(C)	(D)
<b>Panel A. Students and incarceration exposure</b>				
	Individual	School/grade	School	
Mean # students	1.0	192.2	715.2	
Mean # HH incarcerations/year	0.023	4.5	16.7	
<b>Panel B. Estimates from paper</b>				
	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
<b>Panel C. Effect of 1 HH incarceration in school at mean</b>				
	School/grade 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 household incarceration on math and English scores (column (C) of Table 7). Column (B) displays event study estimates of the indirect impacts of household incarceration on math and English scores (column C of Table 8). 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 percent of the raw gradient, which is the estimate in each column divided by the estimate in column (C).

**\*\*\*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 A1: Range of Outcomes under North Carolina Structured Sentencing — Misdemeanors

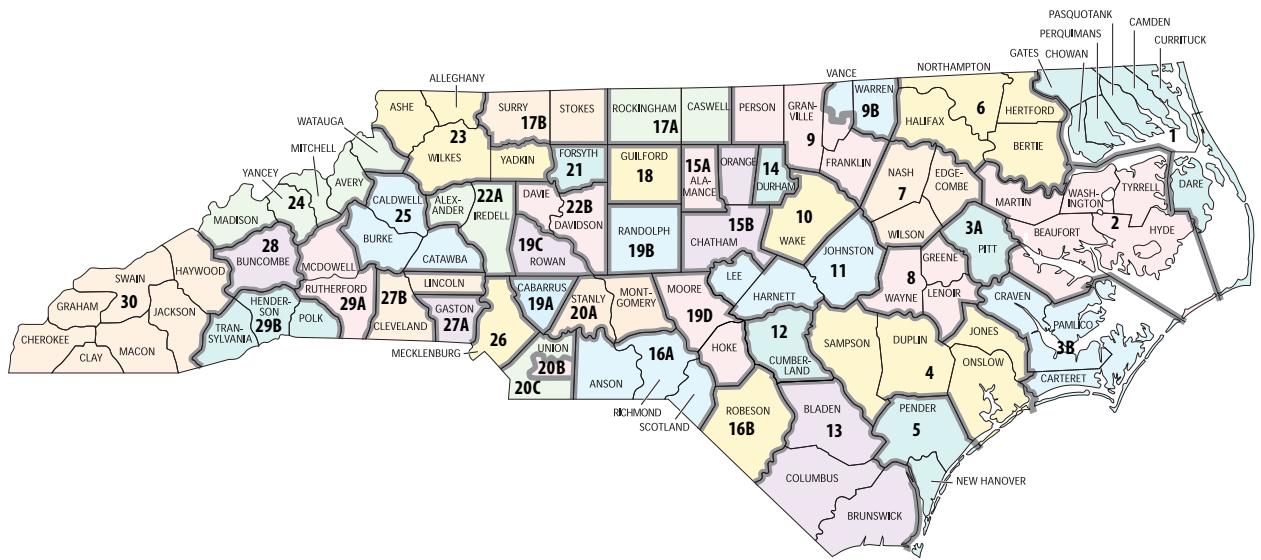
*Notes:* This matrix illustrates the range of judicial discretion for misdemeanor cases in North Carolina under structured sentencing laws during our sample period. This document is taken from the Citizen's Guide to Structured Sentencing, available from the North Carolina Courts at: <https://www.nccourts.gov/documents/publications/citizens-guide-to-structured-sentencing>. 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 \*\*\***

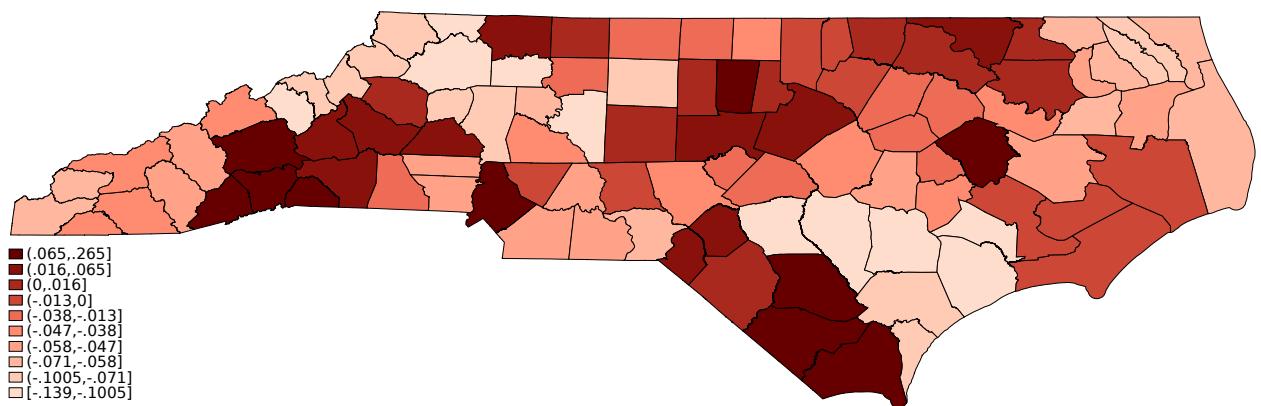
		I 0-1 Pt	II 2-5 Pts	III 6-9 Pts	IV 10-13 Pts	V 14-17 Pts	VI 18+ Pts	
		Death or Life Without Parole						DISPOSITION
		Defendant Under 18 at Time of Offense: Life With or Without Parole						Aggravated Range
		A	A	A	A	A	A	PRESUMPTIVE RANGE
A	B1	A 240 - 300 <b>192 - 240</b>	A 276 - 345 <b>221 - 276</b>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

FIGURE A2: Range of Outcomes under North Carolina Structured Sentencing — Felonies

*Notes:* This matrix illustrates the range of judicial discretion for felony cases in North Carolina under structured sentencing laws during our 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.



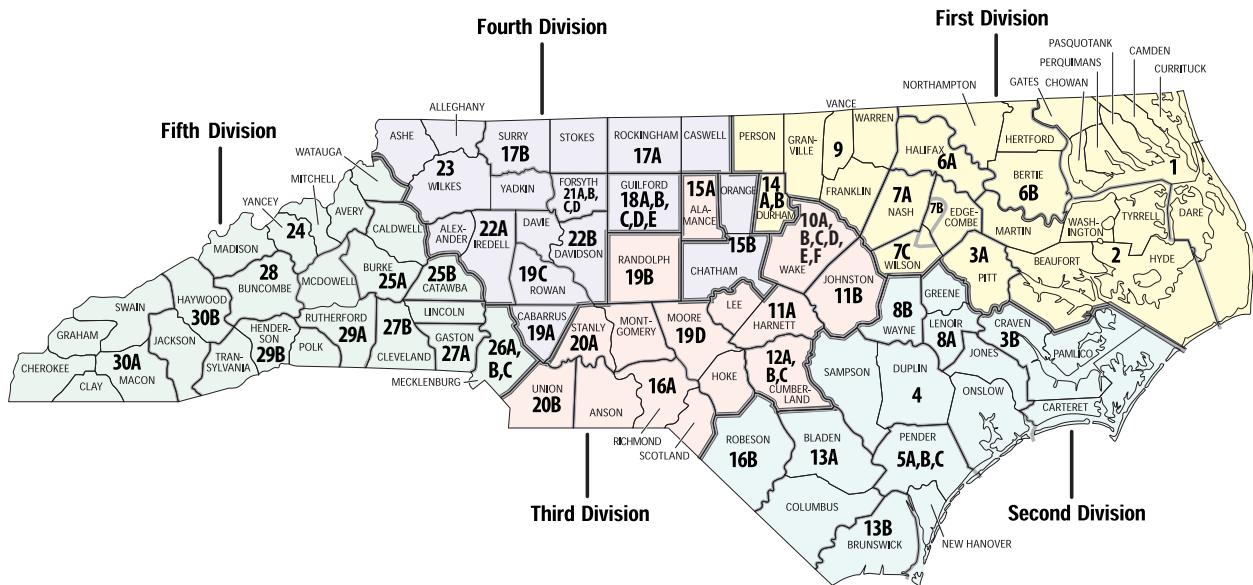
Panel A. District court districts (Effective January 1, 2019)



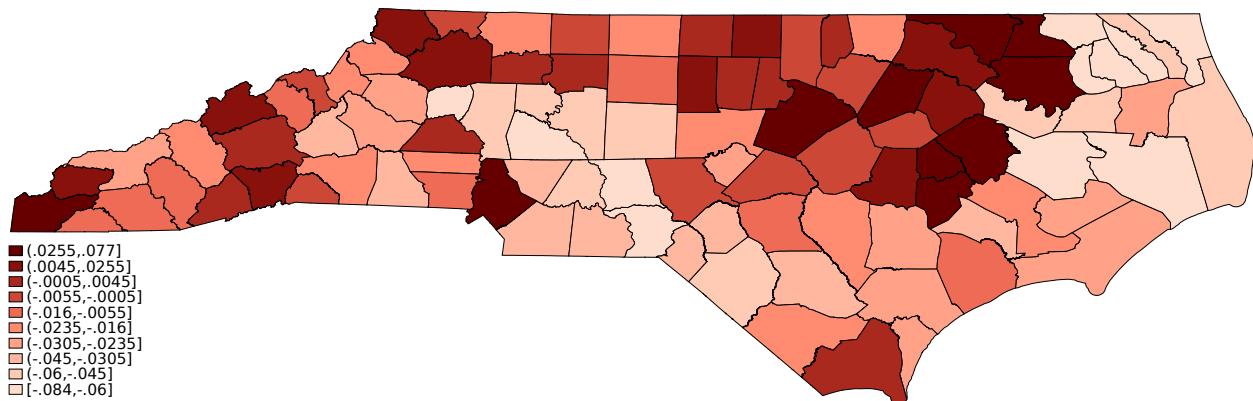
Panel B. Mean stringency for district court judges (2010)

FIGURE A3: 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 A4: 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.

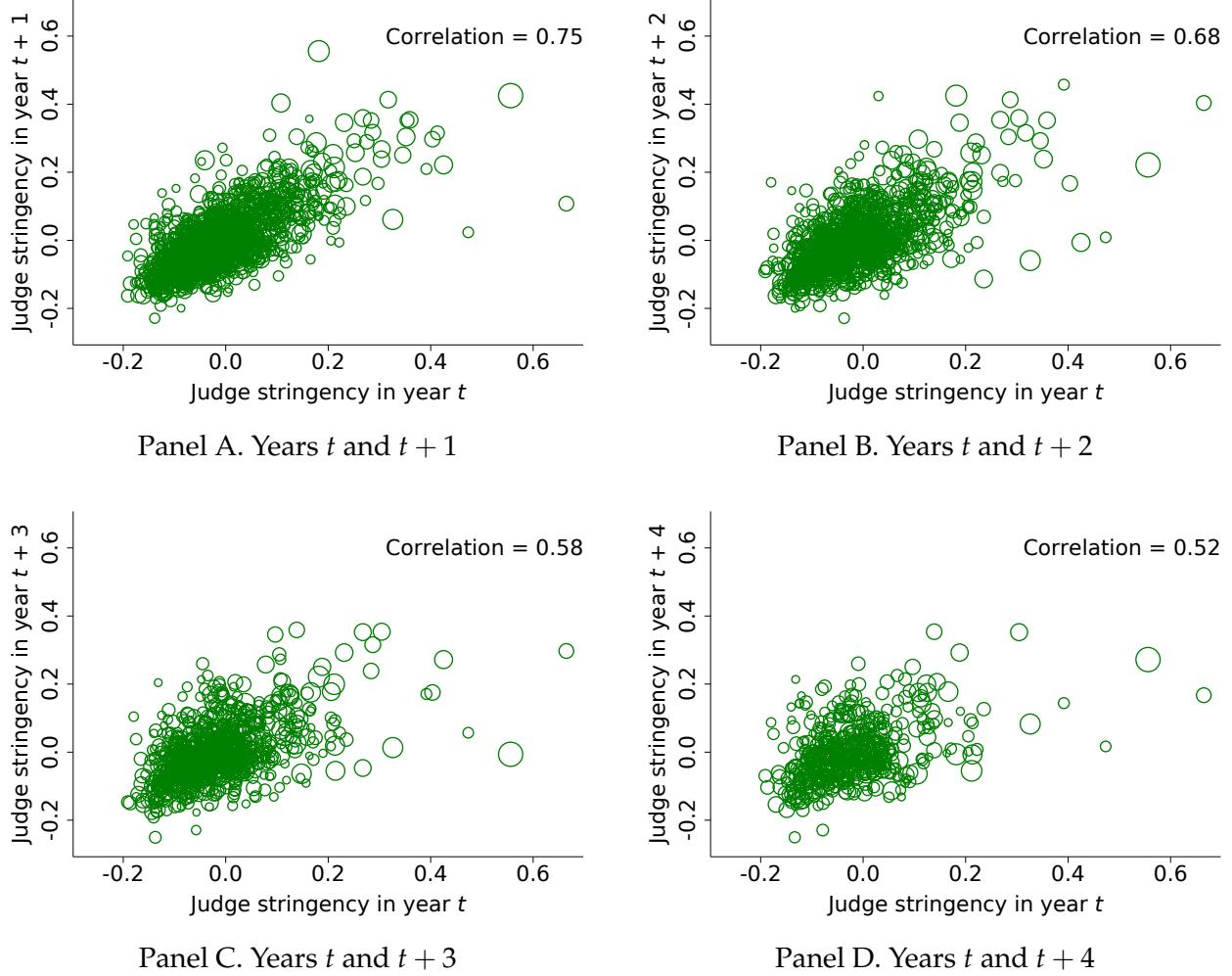


FIGURE A5: Persistence of individual judge stringency

*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.

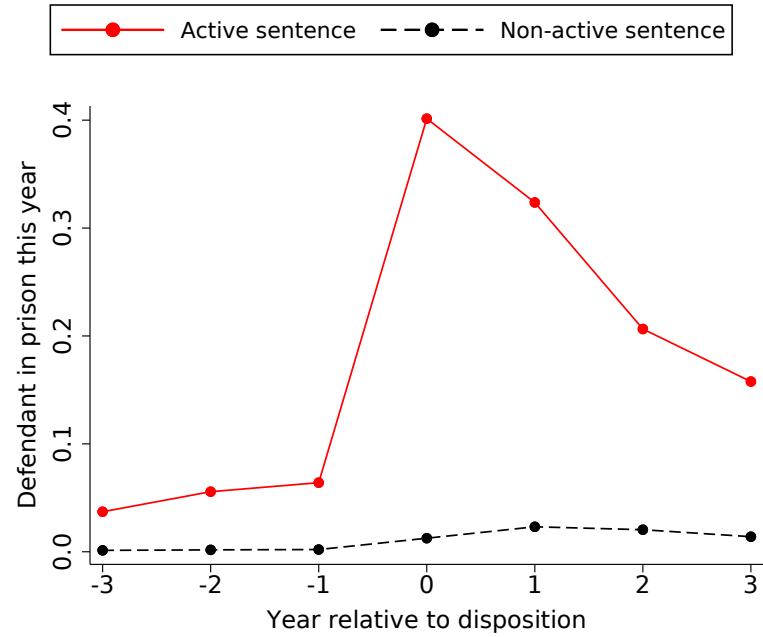


FIGURE A6: Timing of prison spells for defendants

*Notes:* This figure shows the timing of prison spells for defendants who are matched to children in our direct event study sample (column (D) of Table 1). The  $x$ -axis is years relative to disposition. The  $y$ -axis shows the proportion of defendants who appear in prison in each year. The prison records come from the North Carolina Department of Corrections (see Appendix C.2.3), and thus do not include any spells in county jails. The red solid lines depict prison spells for defendants with active sentences, and the black dashed lines show prison spells for defendants with non-active sentences.

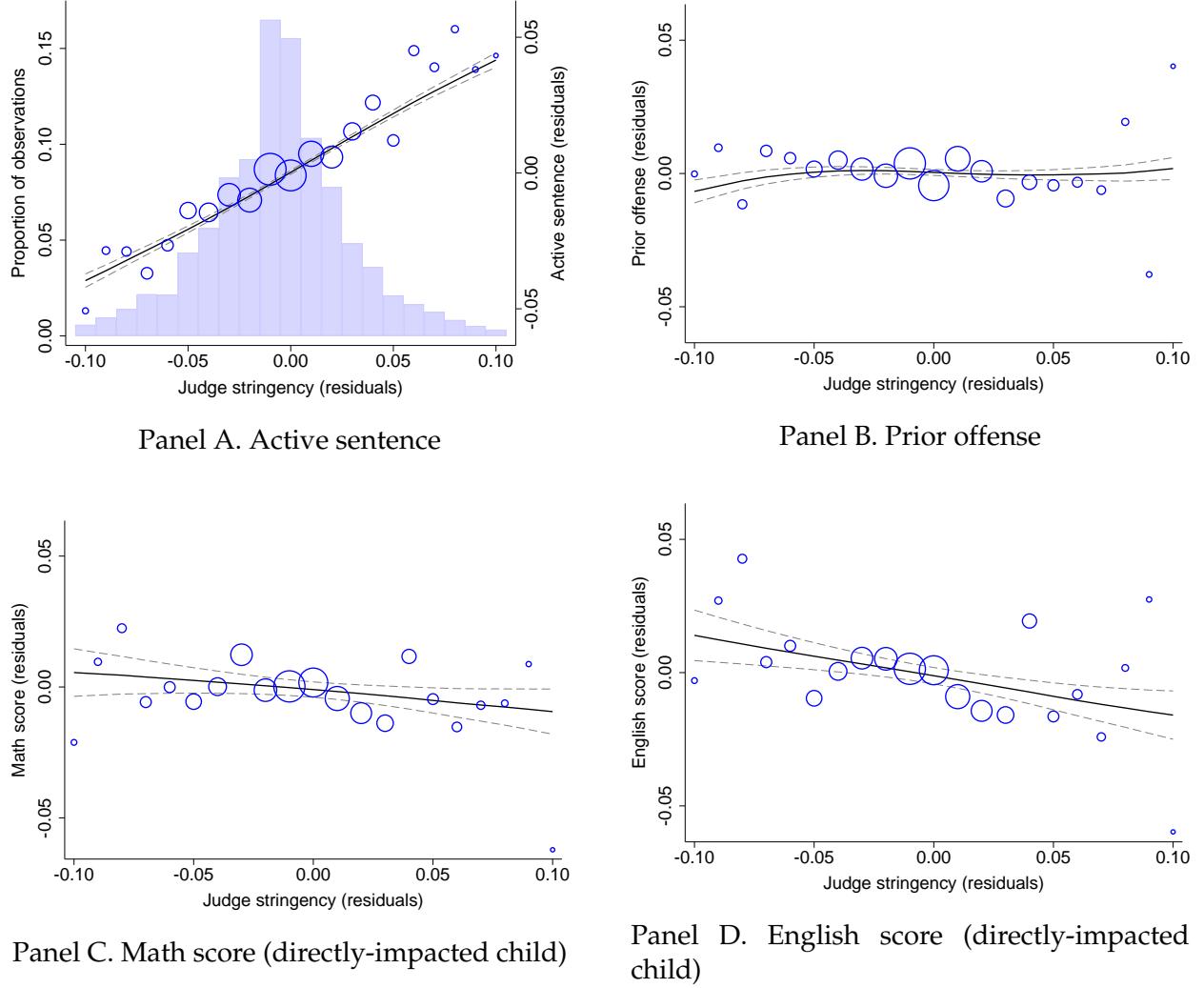
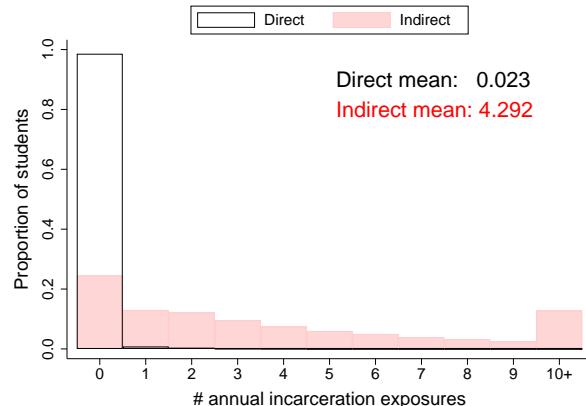
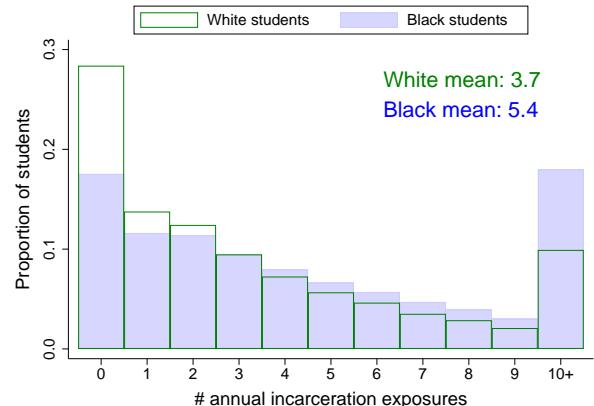


FIGURE A7: Reduced form judge stringency effects

*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 E of Table 1) 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 (two percent of the sample).



Panel A. Direct and indirect exposure



Panel B. Indirect exposure by race

FIGURE A8: 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. 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.

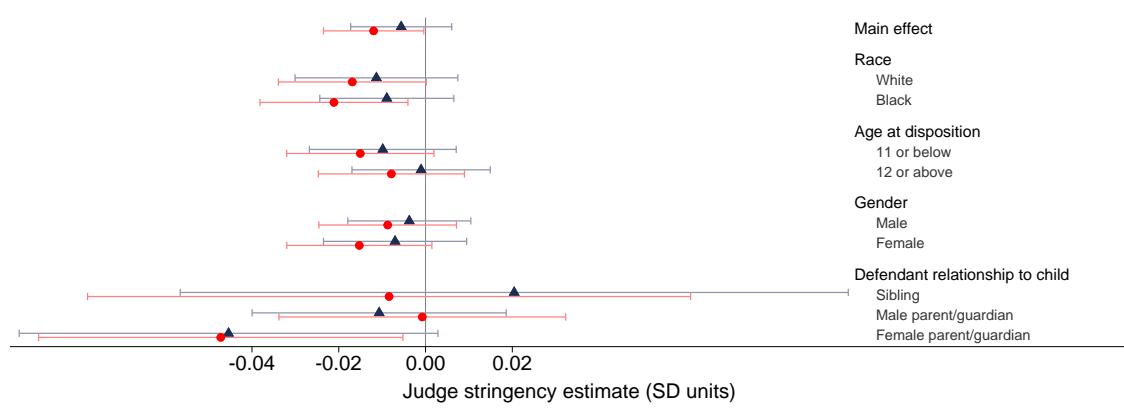
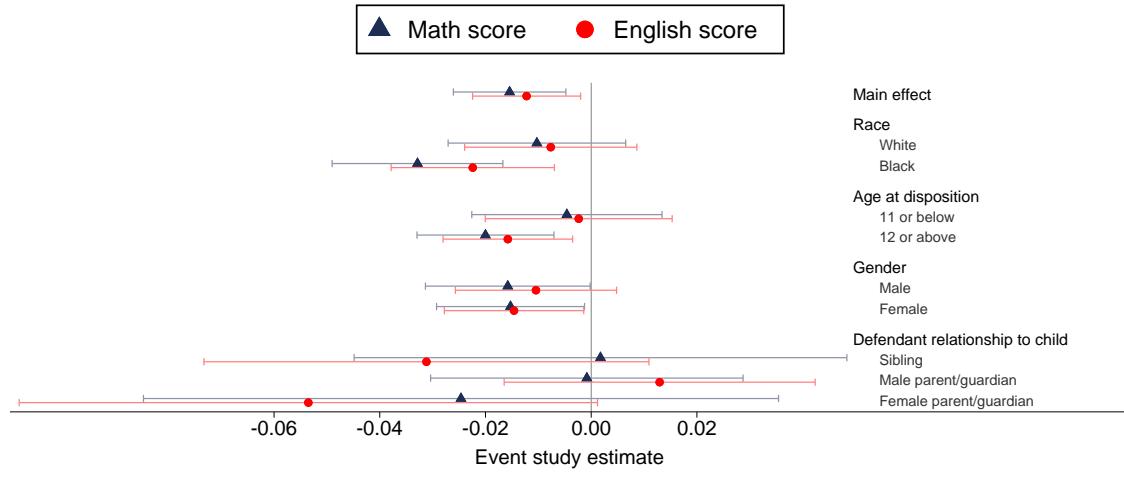


FIGURE A9: Heterogeneity in direct effects of household incarceration on test scores

*Notes:* This figure displays heterogeneity in the direct impacts of household incarceration on student test scores. Panel A presents event study estimates using our direct exposure sample from column (D) in Table 1. Panel B presents reduced-form judge stringency estimates using our direct exposure sample from column (E) in Table 1, 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. We define parent/guardians as defendants who are 20–40 years older than the student. 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.

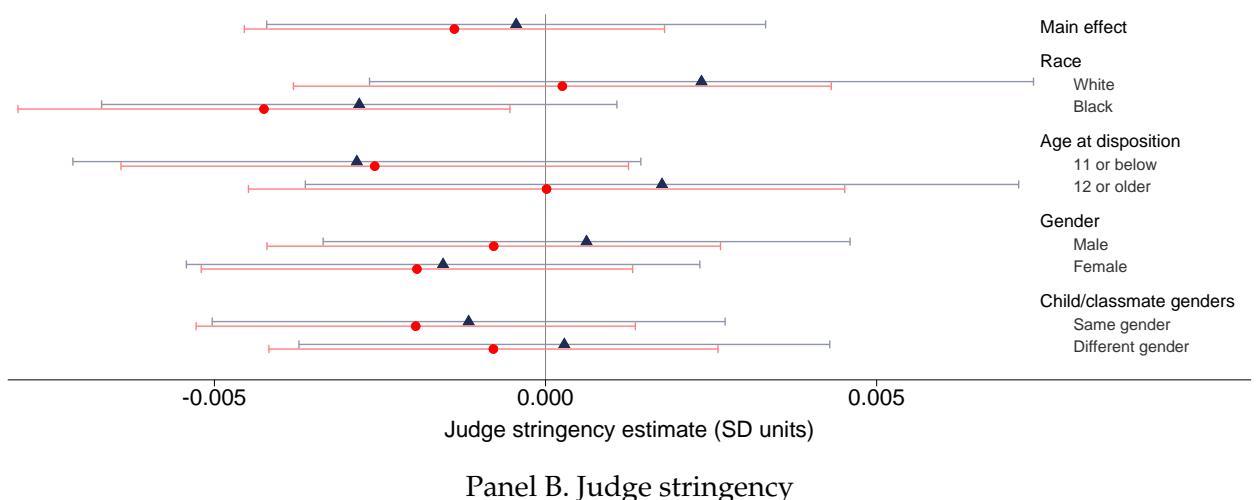
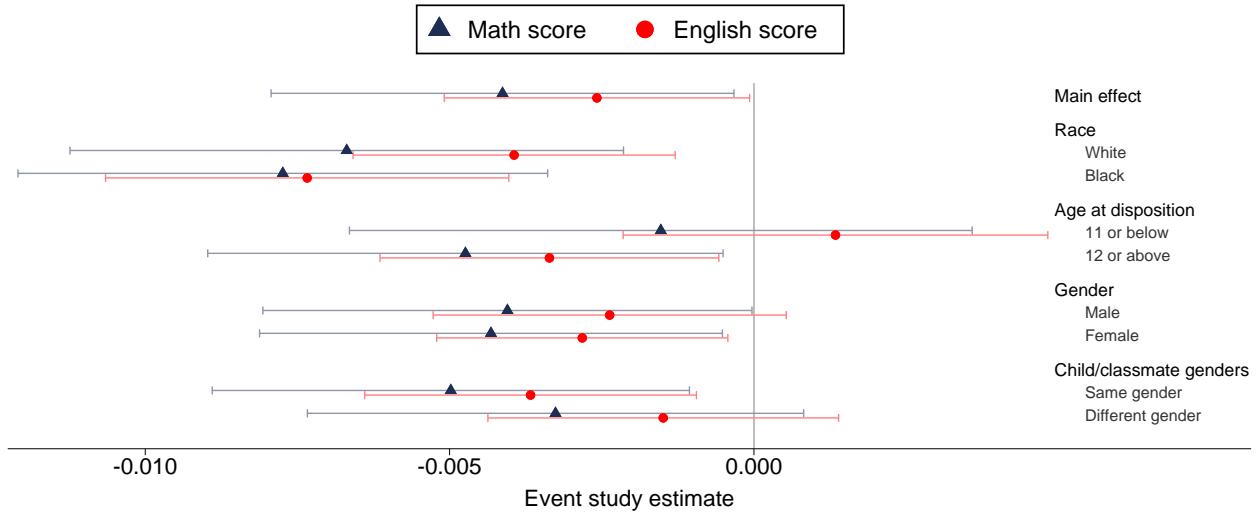


FIGURE A10: Heterogeneity in indirect effects of household incarceration on test scores

*Notes:* This figure displays heterogeneity in the indirect impacts of household incarceration on student test scores. Panel A presents event study estimates using our indirect exposure sample from column (F) in Table 1. Panel B presents reduced-form judge stringency estimates using our direct exposure sample from column (G) in Table 1, 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 8. 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, with the exception of 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.

## B 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.

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.<sup>33</sup> 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.<sup>34</sup>

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 A2. 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 conditions<sup>35</sup>), and 3) active punishment—incarceration in prison or jail. An active sentence necessarily implies judicial

<sup>33</sup> 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>.

<sup>34</sup> 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%20of%20Session\\_Out%20of%20County.pdf](https://benchbook.sog.unc.edu/sites/default/files/pdf/Out%20f%20Term_Out%20of%20Session_Out%20of%20County.pdf).

<sup>35</sup> 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.

incarceration for at least the minimum sentence length.<sup>36</sup> 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.

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 A2. 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.

The District and Superior Court systems differ in several key respects. Figure A3 shows District Court Boundaries, which often follow county boundaries for larger counties. District Court judges are elected for four-year terms, and Superior Court judges serve eight-year terms. While District Court judges stay in their elected district, Superior Court judges rotate across courts within their division according to predetermined Master Schedules. This rotation system generates further variation in caseloads across judges in our sample. Superior Court division boundaries are shown in Figure A4.

---

<sup>36</sup> 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.”

## 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 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.
- **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 9<sup>th</sup> or 10<sup>th</sup> 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 A1–A2). 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 9<sup>th</sup> 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.
- **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 A1–A2). 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).

## 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 (dis-

trict/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 include the date of arrest, arraignment, and disposition, as well as a separate 4-digit offense code for each of these three events.<sup>37</sup> 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 (*crolno*). 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

---

<sup>37</sup> Offense codes are further aggregated into types (Felony, Misdemeanor, and Traffic) and classes (e.g., Felony F, Misdemeanor 2).

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 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.<sup>38</sup>

---

<sup>38</sup> Obtained in November 2020 from: <https://data.census.gov/cedsci/>.

- Latitude and longitude of 2010 Census blocks.<sup>39</sup>
  - Data on child incarceration outcomes by Census tract from the Opportunity Insights project.<sup>40</sup>
  - A crosswalk between 2000 and 2010 Census blocks.<sup>41</sup>
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

---

<sup>39</sup> Obtained in February 2021 from:

<https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2010.html>.

<sup>40</sup> Obtained in February 2021 from: <https://opportunityinsights.org/data/>.

<sup>41</sup> Obtained in November 2020 from: <https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2010-census-block-record-layout.html>.

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.<sup>42</sup>

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 (`crolno`) 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 `crdddt`.

We create a defendant panel in three steps. First, we create a unique ID number for individual defendants.<sup>43</sup> 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

---

<sup>42</sup> Probation sentences have commitment prefixes that begin with numbers.

<sup>43</sup> Creating a defendant ID also allows us to examine defendant recidivism outcomes.

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 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.<sup>44</sup> 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 (`crolno`) 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

---

<sup>44</sup> Specifically, the offense class ordering is Felony A > ⋯ > Felony I > Misdemeanor A1 > ⋯ > Misdemeanor 3 > Traffic A1 > ⋯ > Traffic 3 > Infraction.

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.<sup>45</sup> Cases are almost always 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 A3). Superior court judges rotate to different courts in the same *division*, which are groups of *districts* (see Figure A4).

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

---

<sup>45</sup> The case's county is indicated by the first three digits of the case number crrkey. Case types are defined by the variable crrtyp, which appears in the case records dataset (see Section C.2.1).

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.

#### 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.<sup>46</sup>

Using this 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

---

<sup>46</sup> 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.

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_addrid`) 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.

## 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 (D)–(E) of Table 1). 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 (F)–(G) of Table 1). 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 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 (D)–(E) of Table 1), 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.<sup>47</sup> We exclude defendants that never received a guilty verdict (e.g., if their case was dismissed or if they were found not

---

<sup>47</sup> 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).

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.<sup>48</sup> 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 column (D) of Table 1. 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.<sup>49</sup> We exclude cases that never faced a judge or that never received *any* verdict; these are primarily cases that were dismissed by the

<sup>48</sup> 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.

<sup>49</sup> 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.

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 drop 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 (E) of Table 1. 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 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 (F)–(G) of Table 1) since indirect exposure to incarceration is common (see Appendix Figure A8). 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.