Community Impacts of Mass Incarceration*

Arpit Gupta† Christopher Hansman‡ Evan Riehl§

NYU Stern Imperial College London Cornell University

May 3, 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 student achievement. Our results highlight important consequences of incarceration for access to opportunity within entire communities.

JEL codes: K14, K42, I24, J13

*We thank Jason Baron, Steve Billings (discussant), Kara Bonneau, Ryan Fackler, Ethan Frenchman, David Hoke, Michael Mueller-Smith, Aurélie Ouss, John Rubin, and Jessica Smith for helpful conversations, and conference participants at the Texas Economics of Crime Workshop and NYU Law. 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).

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

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

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

1 Introduction

The unprecedented growth in incarceration since the 1970s has left more than two million adults in jails or prisons in the United States, and underprivileged communities have borne the brunt of this expansion. Mass incarceration disrupts local social norms and networks (Clear, 2008), which may partly explain why heavily exposed neighborhoods have lower intergenerational mobility (Chetty, Hendren, Kline, and Saez, 2014; Chetty, Friedman, Hendren, Jones, and Porter, 2018). Consistent with this notion, Figure 1 displays a large negative relationship between children's achievement in math and the local prevalence of incarceration in North Carolina. However, it is difficult to credibly separate incarceration rates from criminal activity, household sorting, and other aspects of schools and neighborhoods that shape children's opportunities. As a consequence, prior work in economics has primariliy focused on how incarceration affects criminal defendants and their family members, with less attention paid to impacts on the wider community.

Our key contribution is to show that increases in incarceration causally reduce student achievement in the entire community. We do so with an identification strategy that isolates plausibly exogenous variation in local incarceration rates stemming from judicial turnover. Crucially, the effects we find are largely the result of spillovers onto children who *do not* directly experience the incarceration of a parent or household member themselves. While children directly exposed to household incarceration do experience adverse academic and behavioral consequences themselves, they make up a relatively minor portion of the population. Conversely, a much large fraction of students are indirectly exposed to incarcerations through interactions with friends, schoolmates, and neighbors. We document one such spillover channel—behavioral disruptions from directly-impacted children onto their classmates—although many others are possible.

We begin our paper by estimating the impact of community-level incarceration rates on student achievement using a novel dataset and empirical strategy. Our dataset links criminal justice and educational records for the entire state of North Carolina using a merge based on defendant and student addresses. We implement a judge turnover design that exploits year-to-year changes in the composition of judges in each county court. Specifically, we focus on the entry and exit of judges who differ in stringency—their tendency to incarcerate criminal defendants.[1] This strategy

[1] Judges enter and exit as a consequence of elections, retirements, promotions, and cross-county rotations. Our judge turnover design is similar in spirit to the empirical strategy in Chetty, Friedman, and Rockoff (2014) that exploits

builds on work using variation in the stringency of judges assigned to individual defendants (e.g., Mueller-Smith, 2015), but focuses on aggregate variation within counties over time. This turnover has a significant impact on local incarceration: we show that a one standard deviation increase in average stringency leads to a roughly 15–20 percent increase in the county incarceration rate. Further, we show that changes in county-level stringency are unrelated to trends in criminal activity, the demographic characteristics of defendants, and the characteristics of local students.

Our main finding is that increases in community incarceration rates lead to meaningful reductions in the test scores of local students. A one standard deviation increase in county-level stringency reduces student's math scores by 2.5–3.5 percent of a standard deviation, and reduces their English scores by 1.5–2.0 percent of a standard deviation. The effects are persistent, growing in magnitude over the course of at least three years and are nearly identical even when considering only students with no direct household exposure to the criminal justice system. This suggests that the results are largely driven by spillover effects of incarceration onto other children within the community.

The remainder of our paper investigates the mechanisms driving this community-wide effect. We focus primarily on one natural channel: spillovers from children exposed to household incarceration onto their classmates. In order for this channel to operate, we expect to find adverse consequences for children directly exposed to background incarceration effects themselves. However, a number of confounds make it challenging to estimate the academic impacts of incarceration shocks on family members. Students with incarcerated household members likely differ from other students in unobservable ways and may experience time varying shocks alongside an incarceration event (e.g., from arrests and criminal activity).

We use our address-based merge between defendant and student data and two standard empirical strategies to overcome these confounds. The first strategy is an event study that compares student outcomes before and after an incarceration, relative to other students with household members who were convicted of the same crime but not incarcerated. The second is a within-court judicial randomization approach that isolates judge-driven variation in incarceration rates for criminal defendants charged with equivalent crimes. Both the event study and judicial randomization designs perform well in tests for pre-trends and balance on observables.

the arrival and departure of teachers who vary in value added.

We find adverse impacts on academic and behavioral outcomes for directly impacted students using these approaches. Event study estimates suggest that a household incarceration reduces children's math and English scores by 1.0–1.5 percent of a standard deviation. Our estimates are larger when the incarcerated individual appears to be the student's mother, consistent with a role for parental inputs, and larger for Black students. Importantly, we also find negative consequences for a variety of behavioral outcomes, including absences, suspensions, and fighting incidents.

We then show that incarceration reduces the academic performance of the *classmates* of directly-impacted children, and that behavioral disruptions are a likely mechanism for this spillover. We repeat our event study and judge randomization strategies, but consider outcomes for students who share a school and grade level with a directly affected child (and are not directly impacted themselves). Our event studies suggest that this form of indirect exposure reduces math and English scores by 0.4 and 0.3 percent of a standard deviation, respectively. The effects are larger in magnitude in cases where the directly impacted child's misbehavior worsened.

While our spillover estimates are relatively small per student, a large number of children are indirectly exposed to classmates facing household incarceration. As a result, classroom spillovers aggregate to explain a meaningful fraction of the community-level incarceration-achievement gradient. We estimate that causal effects of incarceration 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 relationships—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. Taken collectively, our results suggest that mass incarceration has widespread negative consequences for local access to opportunity.

Our findings do not necessarily imply that incarceration is an unmitigated negative policy. Our study focuses on a particular set of student outcomes, and on the criminal justice system in North Carolina—which, like many in the United States, may do an inadequate job at rehabilitation. It is possible that reintegration based approaches could 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 student achievement. 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. The 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 impacts 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.[2] 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 classroom spillovers as one important mechanism.

A further 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). 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. While these other factors are surely important, our findings suggest that there are widespread consequences of incarceration itself for access to opportunity within communities.

[2] 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).

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 and behavioral outcomes. Other work uses judge randomization designs to estimate the effects of parental incarceration, 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 attainment in Colombia. Perhaps closest to our paper, Norris, Pecenco, and Weaver (2021) find decreases in later life incarceration and improvements in neighborhood quality for directly impacted students in Ohio.

A reconciliation of our results on direct impacts for students with this literature comes in viewing the effects of incarceration as multidimensional. As Murray and Farrington (2008) emphasize, childhood exposure to parental incarceration, while potentially traumatic, may lower children's crime rates by removing negative role models or via deterrence. Indeed, for completeness, we follow Norris et al. (2021) in considering crime rates for children themselves and find similar estimates, suggesting that directly exposed children are less likely to engage in criminal activity. However, our sample permits a more powerful test for the adverse academic and behavioral impacts that we focus on. We interpret the contrasting effects in different domains as indicative of the multidimensional nature of incarceration shocks. A child can simultaneously be traumatised by the incarceration of a parent and steered away from a criminal path.

More fundamentally, however, our central contribution to this literature is to shift the focus 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 county-level approach provides novel evidence on the aggregate impacts of incarceration on children's achievement. The resulting estimates are an order of magnitude too high to be explained through the direct channel alone. Instead, we emphasize that modest spillover effects inside and outside the classroom can aggregate to explain a sizable fraction of the community incarceration-achievement gradient.

2 Data and the Community Achievement-Incarceration Gradient

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

Our linkage, and detailed information on the schools and neighborhoods of each student, allow us to examine educational consequences for students directly exposed to incarceration, as well as spillovers onto the children that they interact with. While past research has considered 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, and to consider differences across communities. By construction, our merge also focuses on criminal defendants associated with children based on address. This enables us to focus on the role of incarceration at the household level, whereas prior literature has typically linked criminal defendants to children by birth records. The incarceration of a household member may be more disruptive to children since many children do not live with both of their parents; in 2010, roughly one in four U.S. children—and more than half of Black children—lived without a father in the household.[3]

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

[3] See: https://www.census.gov/data/tables/time-series/demo/families/children.html.

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 a judges make in sentencing is whether to administer a community punishment (fines or probation), an intermediate punishment (probation with additional restrictions), or an active sentence, which entails incarceration in prison or jail. We use the terms active sentence and incarceration interchangeably throughout the rest of the paper. 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.[4]

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.[5] We observe students' demographic characteristics and their school and grade in each year. Our main academic outcomes are students' scores on state standardized math and English exams. Our measure of math achievement includes scores on both grade 3–8 math tests and an end-of-course high school algebra exam. English scores include performance on grade 3–8 reading tests and an end-of-course high school English I exam. We standardize all scores to have a mean of zero and a standard deviation of one in the population of test takers for each exam cohort. We observe a wide range of behavioral outcomes such as days of absence, suspensions, and fighting incidents. The NCERDC data also include students' geocoded addresses each year,

[4] 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.

[5] 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.

which facilitates our merge with the court data.[6] 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. We sent the NCERDC a list of the addresses that are available in our court records, which include information on street address, city, state, and zip code. The NCERDC linked these variables to the geocoded student address identifier using their confidential information, and then provided us with a crosswalk between addresses in the two datasets. Using this crosswalk, we link students and defendants who live at the same address in the same year.[7] Appendix C.4 provides details on the merge of the ACIS and NCERDC datasets.

Our comparison of linked addresses and individuals suggests that the merge is of high quality. We geocoded addresses in the court data to obtain the Census Block Group of each defendant (this variable also appears in the education data). We find that 97 percent of linked addresses are in the same Census Block Group in cases where this variable is defined in both datasets (see Appendix Table C1). In addition, we find that 84 percent of linked defendants and students have the same race/ethnicity, even before we impose further sample restrictions (see Appendix Table C2).

Table 1 provides summary statistics on the linked education and court data. Column (A) includes all students who attended a North Carolina public school in 2010–2014. Column (B) includes students who were not linked to a defendant in the court data, while column (C) includes the students who were linked. Column (D) includes only those students who were linked to a defendant that received an active sentence in these years. Panel A displays mean demographic characteristics for students, Panel B shows means of our main outcome variables, and Panel C shows characteristics of the criminal defendants for matched subsets.[8] Relative to other public school students, children who are linked to a criminal defendant are more likely to be Black and economically disadvantaged; they also have lower test scores and more misbehavior incidents.

[6] 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.

[7] 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.

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

These patterns are even more pronounced in the subsample of students who were exposed to an active sentence in their household (column (D)).

Our address-based merge has two important caveats. First, we do not observe the relationship between the student and the defendant. As a result, our results reflect the incarceration of any household member. Table 1 shows that the average linked defendant in our data is 33 years old at the time of case disposition, while the average 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 cohabiting. Since student addresses are recorded only once per year, false matches can occur if families move mid-year.[9] 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 analysis samples. Appendix C.5 provides details on the construction of our samples for our analyses in Sections 3–5.

2.3 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 rates and student achievement. Panel A displays a negative and virtually linear relationship between (log) school-level exposure to incarceration and math scores for all North Carolina public school students from 2010–2014. 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.

These figures reflect a 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,

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

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 rates are a powerful predictor of student achievement.

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, job opportunities, crime rates, and 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. The 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)). This indicates that 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 suggest that a causal mechanism may be 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-level incarceration 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 incarceration exposure on children would be to exogenously change 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 intuition behind the identification approach 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 from both arrivals and departures, which stem from judicial elections, mid-term leaves, and pre-scheduled 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 average stringency as:

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

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 $t-1$, ω_{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 B1 and B2). Within each cell, judges have discretion over the prescribed range of potential sentences. Conditioning on structured sentencing cells therefore isolates the variation in active sentencing that is attributable to judicial discretion.

The second concern is that 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 and year t, we define $\mu_{jct}^{-\{t-1,t\}}$ to be the average value of the active sentencing residual for j leaving out any case in county c in years t or $t-1$. Similar to 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 ω_{jct}. In computing the change in county-level stringency between years $t-1$ and t, ΔS_{ct}, we fix the measure of stringency for each judge, $\mu_{jct}^{-\{t-1,t\}}$. As a result, changes in county-level stringency are driven only by changes in the number of cases that judges hear in that county, as captured by the caseloads weights, ω_{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 ω_{jct} to be the fraction of all criminal cases in county c and year t that were heard by judge j. As a result, ΔS_{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 ω_{jct} across years, and we refer to this as our *fixed caseloads* measure. In this approach, ω_{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.[10] 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

[10] We consider judge j to be active in county c if he or she presides over at least 20 cases.

appointment to higher courts), and (iii) rotations across counties. Appendix B provides details on judicial elections and other sources of judge turnover.

Appendix Figures B3 and B4 show the average stringency of district and superior court judges in each county in 2010, along with boundaries for district and superior courts. 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 terms of the probability of assigning an active sentence), and the change from the 10th-90th percentile is 20 percentage points.

3.2 Changes in County Level Stringency Induce Changes in Incarceration Rates

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

$$\Delta Y_{ct} = \beta \cdot \Delta S_{ct} + \gamma_t + \epsilon_{ct} \tag{2}$$

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.[11] 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. γ_t are year fixed effects, and standard errors are clustered at the county level. Across all specifications, we scale ΔS_{ct} so that one unit corresponds to one standard deviation in the baseline distribution of county stringency (0.07).[12] 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. This 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.

[11] Regressions based on Equation 2 are at the county × year level, but we weight observations by the number of individual-level observations used to compute ΔY_{ct}. The one exception to this is when ΔY_{ct} is measured in logs (Panels A–B of Table 3), in which case we estimate unweighted regressions.

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

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 test for balance by showing estimates from Equation 2 using various county level observables. In Panel B, the outcome variables, ΔY_{ct}, measure changes in different types 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)), ΔS_{ct} is not related to changes in the total number of cases, or to the total numbers of cases by offense type (felony, misdemeanor, traffic, or clerk to decide). 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 ΔS_{ct} is unrelated to changes in the average county level race, age, gender or criminal history of defendants. Panel D shows that ΔS_{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.

The results in Panels B–D of Table 3 indicate that the different sources of judge turnover are, at least on average, unrelated to trends in other factors that impact students' educational outcomes. During the period of our data, all judge elections were non-partisan, and so voters may have had little information on which judges were likely to be stringent. Appendix Figure B5 shows that there is essentially no relationship between ΔS_{ct} and changes in county-level Republican vote shares in the 2010 and 2012 elections. Further, many retirements are driven by age, and thus are unlikely to be systematically related to county trends in other outcomes. We examine how different sources of variation in judge turnover impact our main results in subsection 3.6 below.

A key 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 confound in the descriptive results in Section 2, as well as in many interdisciplinary analyses of the community impacts of incarceration.

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

We now present the main results from our judge turnover design, which show that increases in county-level stringency reduce the academic performance of children who live in those counties. Given the robust relationship between ΔS_{ct} and incarceration rates, and the absence of any meaningful relationship with criminal case totals, defendant characteristics, and student demographics, we interpret this as causal evidence that community level incarceration adversely affects 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 ΔY_{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 English scores, and 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 a household member 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 ΔY_{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 ΔY_{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.[13] This represents 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, we modify Equation 2 to consider both pre-trends and persistence in our

[13] This is derived from a simple Wald-style ratio of the impacts of county stringency on math scores (-0.024) and 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 and neighborhoods with the highest exposure.

outcome variables, ΔY_{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, ΔS_{ct}, is defined as the difference between t and $t-1$.

Table 5 shows little evidence of pre-trends in these outcomes, and strong evidence that changes in county stringency have persistent impacts. Columns (A) and (B) show that changes in county stringency at time t do not correlate with trends in active sentences in the pre-period, but they do predict persistent changes in the county active sentence rate at all periods after time t. This holds for both our actual caseloads and fixed caseload measures. Similarly, columns (C)–(F) show that there is little relationship between ΔS_{ct} and test score growth in years prior to t, and 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. Across all specifications, tests of joint significance for both pre-period differences are insignificant while tests of joint significance for all three post-period differences are significant. These results suggest that community-level exposure to incarceration has persistent and accumulative effects on student achievement.

3.6 Decomposing Sources of Variation in County Level Stringency

Our baseline turnover strategy uses 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. We find consistent results across the different primary drivers of turnover.

Panel A of Table 6 shows a variance decomposition of changes in county-level stringency. Our interest is in the variance of ΔS_{ct} (rather than S_{ct}) since this is the variation that drives our analysis. Column (A) shows that the variance of ΔS_{ct} is 0.018^2 in our actual caseloads measure, while column (B) shows that the variance is 0.014^2 in our fixed caseload measure. In other words, 60 percent of the total variation in ΔS_{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 four or eight year terms; 2) mid-term leaves due to retirement or appointment to a higher court; and 3) judges moving to other counties, mostly due to the prescribed rotation of superior court judges. For this we use the same caseload weights, ω_{jct}, as in our fixed caseloads measure, but we compute ΔS_{ct} using only judges who arrive or depart for each reason.[14] 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, ΔS_{ct}, using only the source of variation represented in each column. Panel B shows effects on active sentences and 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 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 reduce student achievement. 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. Appendix Table A2 shows that we find similar results when we restrict to variation from close elections (i.e., those with a winning vote share below 55 percent), and when we use only the stringency of retiring judges (i.e., ignoring the incoming judge).

[14] 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.

4 Mechanisms: Direct Household Exposure

The results in Section 3 indicate that aggregate incarceration rates adversely impact the academic achievement of entire communities. The magnitude of this effect, and the fact that it persists in a sample of students that are not directly linked to court records, indicates that a large fraction of the relationship between incarceration and test scores must be due to spillovers onto a broad set of children within the community. While there are many plausible channels through which such spillovers could operate, we test perhaps the most natural mechanism: contact with directly exposed students within schools. If exposure to the incarceration of a household member impacts the academic performance and misbehavior of a student, we might expect this to lead to disruptions for a large number his or her classmates.

We highlight this channel in two steps. In this section, we show a key prerequisite: that direct exposure to incarceration within the household adversely impacts student achievement and behavior. In the next section, we show that these direct impacts spill over onto classmates.

4.1 Empirical Specifications

OLS Estimation

Our object of interest in this section is the causal impact of direct exposure to an active sentence within the household. A simple approach to recovering this effect is a cross-sectional regression comparing 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 student i in year t, our OLS specification is:

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

Here, y_{it} represents a student educational outcome 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.[15] $\theta_{oc\tau(i)}$ is a fixed effect for offense class

[15] Note that many of our variables, including Active Sentence$_i$, are characteristics or 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

at arrest (o) $\times$ court (c) $\times$ disposition year (τ), 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.[16] For these regressions, we consider all students matched to court records with the exception of (i) those linked to low-level traffic offenses, (ii) likely false matches. We also drop cases missing information on a judge (or matched to a judge who appears rarely in the data) and consider only the most serious charge associated with each student. Appendix C.5 provides more specific details on the construction of our sample, and Appendix Table A3 shows sample summary statistics.

There are two main endogeneity concerns with this cross-sectional OLS regression. First, because it compares convicted defendants to those who are charged but not convicted, it risks conflating the traumatic impacts of the offense itself (or the circumstances precipitating the offense) with the effects of incarceration. Second, incarceration status is likely correlated with unobservable defendant or student characteristics that impact student achievement. To address these issues, we adopt two complementary strategies from the literature: an event study approach and an individual level judicial stringency approach.

Event Study Approach

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

Our event study specification considers outcome y_{it} for student i in calendar year t:

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

one case and one defendant, and so we use i as the subscript for these variables to reduce notation.

[16] 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.

η_i is an individual fixed effect and $\lambda_{og\tau(i)t}$ is a granular set of offense class (o) $\times$ academic cohort (g) $\times$ disposition year (τ) $\times$ calendar year (t) fixed effects.[17] The regressions include all years t in which we observe student outcomes, and our variable of interest is the interaction between Active Sentence$_i$, and an indicator for years of or after the case disposition, $\mathbb{1}\{t \geq \tau(i)\}$. Intuitively, the fixed effects restrict identification to students who are in the same graduation cohort, and whose household members were convicted in the same year with similar offenses. Our coefficient of interest, δ, 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)).

This event study requires a slightly different sample relative to our OLS regressions, namely (i) we include observations prior to the disposition year, (ii) we restrict to students matched to *convicted* defendants, and (iii) we select the most serious linked conviction (rather than charge) for each student. Appendix C.5 provides details on the construction of this sample, and Appendix Table A3 shows summary statistics.

Judicial Stringency Strategy

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

We estimate a reduced form specification that mimics equation 3, but replaces Active Sentence$_i$

[17] 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.

with $\mu_{j(i)}$, the stringency of the judge in question:[18]

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

The coefficient of interest, γ, measures the impact of a household member being assigned to a more stringent judge on student outcomes. Because the likelihood of an active sentence increases for those assigned to more stringent judges, this provides a reduced form estimate of the causal impact of incarceration. We also estimate an IV specification in which equation 5 serves as a first stage (with Active Sentence$_i$ on the left-hand side), and Equation 3 serves as a second stage.[19] In both specifications, we use the same sample as in our baseline OLS regressions.

The key identification assumption is that judge stringency, $\mu_{j(i)}$, is unrelated to students' potential outcomes, conditional on controls. The assignment of cases to judges in North Carolina, while not explicitly randomized, is quasi-random; this reflects pre-determined assignments of judges to courtrooms at the discretion of the senior judge in consultation with county clerks. The frequent movement of judges across counties and courtrooms, in both district and superior courts, ensures that judges oversee a variety of offense types. Indeed, balance tests support the assumption that stringency is effectively randomly assigned. $\mu_{j(i)}$ is not correlated with characteristics of the defendant, offense, or cohabitating children, but is strongly related to the probability of receiving an active sentence (see panel A of Table 7). Appendix B discusses the North Carolina judicial process in more detail and presents these balance tests.

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

[19] Under standard independence, exclusion, and monotonicity assumptions, our IV approach recovers 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 monotoncity (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, γ, in our discussion below, but include IV estimates for comparison with related work.

4.2 Consequences of Direct Household Exposure to Incarceration for Students

Across our empirical strategies, we find consistent evidence that direct exposure to incarceration within the household has adverse impacts on student academic performance and behavior. We discuss these in turn below.

Academic Performance

Panel B of Table 7 shows adverse consequences of household incarceration on math and English test scores. Our cross-sectional OLS approach suggests that exposure to an active sentence is associated with a reduction in math test scores of 3.4 percent of a standard deviation, and in English scores of 1.4 percent of a standard deviation. The former is significant at the 1 percent level. Our event study approach—which compares exposure to incarceration among those exposed to a conviction—finds consistent but slightly smaller impacts. Exposure to an active sentence reduces math scores by 1.5 percent of a standard deviation and English scores by 1.2 percent, and both effects are statistically significant. Our reduced form judicial stringency approach, although scaled in different units, finds similar results. A one standard deviation increase in judicial stringency reduces Math and English scores by 0.6 and 1.2 percent of a standard deviation, respectively (although only the latter is statistically significant at the 5 percent level). Point estimates for the IV specifications are qualitatively similar, but substantially larger. Together, our approaches suggest that direct exposure to incarceration has meaningful negative impacts on academic outcomes.

Attendance and Misbehavior

Panel C of Table 7 shows that direct exposure to household incarceration leads to increases in absences and misbehavior incidents. Our cross-sectional OLS estimates indicate that direct exposure is associated with 0.4 additional days absent, a roughly 6 percent increase in the probability of suspension, 0.15 additional suspension days, and over a 6 percent increase in the probability of being disciplined for a fighting incident (these effects are all statistically significant). Our event study estimates are extremely similar to these OLS specifications, with minor variation in magnitudes across outcomes. Finally, 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 reduced form estimates are generally smaller in magnitude than those in our event study (and are scaled in different units), but they are still economically meaningful given the wide variation in judge stringency.

Graphical Evidence From Event Studies

To show evidence supporting the identification assumption of parallel trends for our event study approach, we estimate a modified version of equation 4 that replaces the $\mathbb{1}\{t \geq \tau(i)\}$ term with dummies for years relative to the defendant's disposition (omitting the year prior to disposition).[20] Figure 2 shows the coefficients on these interaction terms from three years before through three years after conviction. In Panels A–D, we see little evidence of differential pre-trends between treated and control students in the years leading up to conviction for academic and behavioral outcomes. After 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.

Heterogeneity

We examine heterogeneity in student outcomes in Appendix Figure A1. Panel A 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 also find 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. We explore heterogeneity in the judge stringency effects on test scores in Panel B. Our results are

[20] Specifically, Figure 2 plots δ_l coefficients from the specification:

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

where 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.

largely consistent with our event study analysis; in particular, incarcerations of likely mothers are associated with large and negative impacts on children.

Robustness

Panel A of Appendix Table A4 shows robustness tests for our event study approach. For these we layer in granular controls for the defendant's offense (o) to examine the importance of other time-varying factors that may be related to the severity of the criminal activity or arrest. Column (A) shows results with *no* offense controls; column (B) controls for offense class the time of arrest; column (C) controls for offense class the time of conviction (our benchmark specification); column (D) controls for four-digit offense codes at conviction; and column (E) controls for four-digit offense codes plus a time trend in the defendant's prior points. The magnitudes of our event study estimates decline slightly across columns, but they remain significant and economically meaningful for most outcomes even with the inclusion of detailed offense characteristics. This suggests that our event study results are driven by the defendant's incarceration rather than the factors that influence the sentencing outcome.

Panel A of Appendix Table A5 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 changed with felonies (column (E)). Appendix Figure A2 displays non-parametric versions of our judge stringency estimates, presenting evidence that stringency predicts receipt of an active sentence (panel A), is uncorrelated with prior offenses (panel B), and predicts math and reading scores (panels C and D).

4.3 Relationship with Literature on Direct Impacts

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 stringency designs, and vary in whether they find positive or negative impacts on children. To understand how our results fit in with this literature, we draw on Murray and Farrington (2008), who emphasize the multidimensional impacts of incarceration, including (1) direct psychological strains, (2) removing criminal role models (which might lower children's criminality rates), and (3) other shocks to human capital that operate through family resources. In other words, the incarceration of a family member may have adverse consequences in some domains while leading to improved outcomes in others.

Our primary contribution to this literature comes in highlighting important adverse consequence of direct exposure. This includes increases in school disciplinary events, which can be viewed as evidence of the psychological strains of household incarceration. In the next section, we explore the role these behavioral disruptions play in generating spillovers onto other students within the community.

More generally, we view our findings as consistent with the most similar studies in a US context. Given the multidimensional impacts of incarceration, we expect that the adverse effects we estimate likely co-exist with the positive impacts emphasized, e.g., in Norris et al. (2021), which uses a judicial stringency design in Ohio. To underscore this point, Panel D of Table 7 considers two key outcomes from that study: neighborhood socioeconomic status percentile and the probability that the directly impacted student themselves has a criminal charge.[21] Of course, this comparison is limited by the fact that we only observe outcomes until the child is in grade 12, as opposed to outcomes in young adulthood. Still, despite this caveat, point estimates from our event study and judicial stringency estimates are in line with the conclusions of this prior work. We find positive effects on neighborhood quality for directly impacted students, suggesting that families move to better neighborhoods following a household incarceration. We find negative point estimates for criminal charges, suggesting that these students are at least no more likely to commit offenses themselves (although child offenses are infrequent given our data structure). Finally, our estimates on the test score impacts of incarceration are relatively modest, and thus lie within the confidence intervals of the estimates in much of this work (including Norris et al., 2021). This suggests that our design is powered to pick up moderate, but still meaningful, impacts of incarceration on academic achievement.

Thus the composite nature of our findings is both consistent with prior literature and highlights the multidimensional consequences of incarceration. For example, direct exposure to incarceration may destabilize a household and negatively impact student achievement, while at the

[21] See Appendix C.1 for details on how we define these variables in our data.

same time providing first-hand interactions with the justice system that act as a deterrent to future criminal behavior. This interpretation warns against viewing incarceration as uniformly good or bad, but instead points to the more nuanced conclusion that it can have varied impacts on children. Furthermore, the key focus of our study is on the broader consequences of incarceration for entire communities. The next section highlights one specific channel, spillovers within classrooms that result from behavioral disruptions.

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 aggregate achievement-incarceration gradient, we next examine one potential channel for spillovers onto indirectly exposed children: interactions with directly impacted children within the classroom. We show that this channel is a meaningful driver of the community level relationship.

5.1 Empirical Specifications

To examine indirect impacts, we use the same empirical strategies as in Section 4, but we focus on students $k \in \mathcal{K}(i)$ who were classmates with some child i in our direct exposure samples. For each directly-exposed child i, we define the set of classmates, $\mathcal{K}(i)$, as the students who were in the same school and grade as child i in the year of the defendant's disposition. We refer to these students as the "classmates" of child i because they tend to move through the school system together, and thus frequently share classrooms. Appendix C.5 provides details on our indirect exposure samples, and Appendix Table A3 displays summary statistics.

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:

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

This specification is nearly identical to Equation 4, but observations are at the *kit* level and we include a fixed effect, η_{ki}, for each classmate $\times$ child pair. All other covariates are still defined by the defendant linked to child i. Similarly, our OLS and judge stringency specifications are generalizations of Equations 3 and 5 to the *kit* level.[22]

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.

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 on 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

[22] Specifically, our OLS and reduced-form judge stringency specifications for indirect impacts are:

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

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

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

in our event study specification. The point estimates imply that the incarceration of a child's household member lowers their classmates' math and English scores by 0.4 and 0.3 percent of a standard deviation, respectively. 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 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 underpowered 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.[23] 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, and negative indirect impacts on test scores following the assignment of a stricter judge.

We demonstrate balance and robustness for our estimates of indirect effects in a set of tests that is comparable to our analysis for directly affected students. In Panel B of Appendix Tables A4 and A5, 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 B1 shows that judge stringency is not significantly related to characteristics of classmates in our indirect exposure sample (column (D)).

5.3 Channels of Indirect Impacts

Given the evidence that direct exposure to incarceration increases misbehavior (shown in Section 4), we hypothesize that this misbehavior may be a key mechanism driving the indirect effects shown here. Following Lazear (2001)'s seminal model of classroom disruptions, there is now a

[23] 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.

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, Paserman, 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 ratio of indirect-to-direct impacts of incarceration 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 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 figure, 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.[24] 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.[25] Red diamonds show average values for ventiles of the x-axis variable, 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 *particularly* in cases where the directly-impacted child's misbehavior increased.

Appendix Table A6 provides tests that formalize the results in Figure 4. To construct these

[24] To account for the fact that suspension rates increase with age, as well as any effects of the crime or 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 × academic cohort × disposition year) whose household member did *not* receive an active sentence.

[25] 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.

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 6), 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 children whose misbehavior increased as a direct result of exposure to incarceration. Panel A of Appendix Table A6 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 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.[26]

6 Discussion

In sections 4–5 we show evidence for one channel underlying the community level relationship between achievement and incarceration: direct impacts on children within the household that spill-over onto their classmates. In this section, as a final step, we quantify the importance of this channel in explaining the aggregate gradient shown in Figure 1. Figure 5 shows a graphical representation of the exercise, with supporting calculations provided in Appendix Table A7. 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

[26] Appendix Figure A3 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.

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 (column (C) in Table 2).[27]

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

By contrast, the indirect channel can explain 9 percent of the raw achievement-incarceration gradient in math. This greater magnitude is the result of the the relative frequency with which students are indirectly exposed to household incarcerations through their classmates. In our data, the typical child is indirectly exposed to 4.3 household incarcerations per year (see Appendix Figure A4). Over 75 percent of public school children in North Carolina have at least one classmate with a household incarceration in a typical year, and over 15 percent of children experience ten or more indirect exposures.[28] Thus, while the indirect effects on math scores are small in magnitude (0.4 percent of a standard deviation), they aggregate to explain a large fraction of the average impact. Appendix Table A7 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, which is likely only one of many causal channels that 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 mechanisms underlying indirect community level impacts.

[27] For this decomposition we assume constant effects at the mean level of incarceration exposure, and we ignore any dynamic effects. See Appendix Table A7 for details.

[28] 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.

7 Conclusion

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

We then highlight an important mechanism underlying this result: spillovers within the classroom. We show that direct exposure to the incarceration of a household member negatively impacts a student's test scores and behavior, which in turn has adverse consequences for the academic achievement of their classmates. The impacts on achievement are persistent, lasting at least three years at both the individual and aggregate levels. We point towards behavioral disruptions as a likely source of classroom-level spillovers. Though small in magnitude for any individual child, these spillovers account for a large share of the overall gradient between achievement and community incarceration rates. 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 incarcerated-individuals and their family members, our paper advances the literature with a research design that allows us to establish the causal consequences of community-level incarceration rates. Our findings point to the value of criminal justice reform that takes into account the broader consequences of mass incarceration. Incorporating the spillover effects of incarceration into the design of policy 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. (2008). The effects of high imprisonment rates on communities. *Crime and justice 37*(1), 97–132.

Clear, T. R., D. R. Rose, E. Waring, and K. Scully (2003). Coercive mobility and crime: A preliminary examination of concentrated incarceration and social disorganization. *Justice Quarterly 20*(1), 33–64.

Dobbie, W., H. Grönqvist, S. Niknami, M. Palme, and M. Priks (2018). The intergenerational effects of parental incarceration. Technical report, National Bureau of Economic Research.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education finance and policy 2*(4), 376–394.

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.

Murray, J. and D. P. Farrington (2008). The effects of parental imprisonment on children. *Crime and justice 37*(1), 133–206.

Neidell, M. and J. Waldfogel (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics 92*(3), 562–576.

Norris, S., M. Pecenco, and J. Weaver (2021, September). The effects of parental and sibling incarceration: Evidence from ohio. *American Economic Review 111*(9), 2926–63.

Pattillo, M., B. Western, and D. Weiman (2004). *Imprisoning America: The social effects of mass incarceration*. Russell Sage Foundation.

Rose, D. R. and T. R. Clear (1998). Incarceration, social capital, and crime: Implications for social disorganization theory. *Criminology 36*(3), 441–480.

Spainhour, W. and S. Katzenelson (2009). Structured sentencing training and reference manual. *Raleigh, NC: North Carolina Sentencing and Policy Advisor Commission*.

Thomas, A. and I. Sawhill (2005). For love and money? the impact of family structure on family income. *The Future of Children*, 57–74.

Wildeman, C. (2010). Paternal incarceration and children's physically aggressive behaviors: Evidence from the fragile families and child wellbeing study. *Social Forces 89*(1), 285–309.

Wildeman, C. and K. Turney (2014). Positive, negative, or null? the effects of maternal incarceration on children's behavioral problems. *Demography 51*(3), 1041–1068.

Tables

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

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

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

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

TABLE 2: OLS 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
Panel A. School exposure: Linear effect of 1 log point in HH incarcerations/student					
Math score	-0.205^{***}	-0.104^{***}	-0.104^{***}	-0.102^{***}	-0.098^{***}
	(0.009)	(0.007)	(0.006)	(0.006)	(0.006)
English score	-0.208^{***}	-0.098^{***}	-0.094^{***}	-0.092^{***}	-0.085^{***}
	(0.006)	(0.005)	(0.004)	(0.004)	(0.005)
N (Math scores)	3,498,907	3,498,900	3,498,900	3,320,477	1,697,216
Panel B. Neighborhood exposure: Linear effect of 1 log point in incarcerations/resident					
Math score	-0.276^{***}	-0.128^{***}	-0.120^{***}	-0.119^{***}	-0.121^{***}
	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)
English score	-0.278^{***}	-0.127^{***}	-0.117^{***}	-0.116^{***}	-0.113^{***}
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
N (Math scores)	3,271,929	3,271,922	3,271,922	3,101,129	1,548,345
Included fixed effects:					
Demographics		Y	Y	Y	Y
County & year			Y	Y	Y

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

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

TABLE 3: County stringency balance tests and effects on active sentencing

	(A)	(B)		(C)	
		Coefficient on Δ county stringency (SD units)			
		Actual caseloads		Fixed caseloads	
Dependent variable	2010 mean (levels)	Coef	(SE)	Coef	(SE)
Panel A. Active sentences					
Δ log # active sentences	622	0.208***	(0.067)	0.166**	(0.064)
F statistic		9.7		6.6	
Panel B. Criminal case totals					
Δ log # total cases	6,802	0.006	(0.013)	0.023*	(0.013)
Δ log # felony cases	904	0.008	(0.030)	0.020	(0.037)
Δ log # misdemeanor cases	3,279	−0.004	(0.018)	0.001	(0.015)
Δ log # traffic cases	2,379	0.015	(0.021)	0.049	(0.031)
Δ log # clerk to decide cases	240	0.041	(0.078)	0.040	(0.108)
F statistic: All coefficients zero		0.3		1.5	
Panel C. Defendant characteristics					
Δ proportion male	0.704	−0.001	(0.002)	0.004	(0.003)
Δ mean age	32.420	−0.051	(0.047)	−0.003	(0.068)
Δ proportion white	0.476	0.003	(0.006)	0.004	(0.007)
Δ proportion Black	0.433	−0.003	(0.007)	−0.003	(0.007)
Δ proportion w/ multiple offenses	0.432	−0.007*	(0.004)	−0.005	(0.004)
Δ proportion w/ prior offense	0.505	−0.001	(0.005)	−0.001	(0.005)
Δ proportion w/ prior active sentence	0.072	0.005	(0.005)	0.003	(0.007)
F statistic: All coefficients zero		1.0		1.2	
Panel D. Student characteristics					
Δ proportion male	0.507	−0.000	(0.001)	0.000	(0.001)
Δ mean age	11.764	0.005	(0.008)	−0.001	(0.011)
Δ proportion white	0.541	−0.002	(0.001)	−0.002*	(0.001)
Δ proportion Black	0.270	0.003**	(0.001)	0.003	(0.002)
Δ proportion economically disadvantaged	0.492	0.001	(0.002)	0.002	(0.002)
F statistic: All coefficients zero		1.5		1.0	
N (# county/years)	100	400		400	
# criminal cases	680,192	3,153,479		3,153,479	
# students × years	1,457,836	7,422,413		7,422,413	

Notes: This table examines how changes in our county-level measure of judge stringency are related to changes in active sentences, criminal case totals, defendant characteristics, and student characteristics. We estimate Equation 2, $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$. ΔS_{ct} is the change in county stringency using both actual caseloads (column B) and fixed caseloads (column C), and γ_t are year fixed effects. Our outcomes, ΔY_{ct}, measure changes in four sets of variables. In Panel A, the outcome variable is changes in 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

	(A)	(B)	(C)	(D)	(E)
		Coefficient on Δ county stringency (SD units)			
		All students		Students not linked to court records	
Dependent variable	2010 mean (levels)	Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads
Panel A. Academic performance					
Δ mean Math score	−0.031	-0.024^{***}	-0.036^{**}	-0.027^{***}	-0.045^{**}
		(0.008)	(0.018)	(0.007)	(0.017)
Δ mean English score	0.001	-0.014^{**}	-0.020^{*}	-0.015^{**}	-0.026^{***}
		(0.007)	(0.010)	(0.006)	(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 Δ Math score	0.015	−0.025	-0.025^{*}	-0.028^{*}	-0.032^{**}
		(0.015)	(0.015)	(0.016)	(0.016)
Mean individual Δ Reading score	0.008	−0.013	-0.018^{**}	−0.013	-0.021^{**}
		(0.010)	(0.009)	(0.010)	(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: $\Delta Y_{ct} = \beta \Delta S_{ct} + \gamma_t + \epsilon_{ct}$. γ_t are year fixed effects. ΔS_{ct} corresponds to our main treatment variable, estimated as in equation 1 and discussed in section 3.1, based on the change in local judicial stringency resulting from judicial arrivals/departures: $\Delta S_{ct} = \sum_j (\omega_{jct} - \omega_{jct-1}) \mu_{jct}^{-\{t-1,t\}}$. Columns (B) and (D) show results using the actual caseload weights ω 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 ΔS_{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

	(A)	(B)	(C)	(D)	(E)	(F)
	Dep. variable: Log # active sentences		Dep. variable: Mean Math score		Dep. variable: Mean English score	
Timing of dependent variable	Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads	Actual caseloads	Fixed caseloads
$\Delta Y_{t-3,t-1}$	0.031	-0.012	0.017	0.028*	0.005	0.012
	(0.039)	(0.070)	(0.013)	(0.016)	(0.010)	(0.011)
$\Delta Y_{t-2,t-1}$	0.055	0.021	0.009	0.016	0.005	0.001
	(0.056)	(0.073)	(0.009)	(0.010)	(0.008)	(0.008)
$\Delta Y_{t,t-1}$	0.208***	0.166**	-0.024^{***}	-0.036^{**}	-0.014^{**}	-0.020^{*}
	(0.067)	(0.064)	(0.008)	(0.018)	(0.007)	(0.010)
$\Delta Y_{t+1,t-1}$	0.261***	0.204**	-0.051^{***}	-0.061^{***}	-0.029^{***}	-0.033^{**}
	(0.072)	(0.082)	(0.015)	(0.022)	(0.009)	(0.015)
$\Delta Y_{t+2,t-1}$	0.264***	0.208***	-0.040^{*}	-0.055^{**}	-0.027^{**}	-0.037^{***}
	(0.074)	(0.072)	(0.022)	(0.021)	(0.012)	(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
F-statistics:						
All prior coefs zero	0.6	0.1	0.9	1.7	0.2	0.7
All post coefs 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, ΔS_{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

	(A)	(B)	(C)	(D)	(E)
			Sources of variation in fixed caseloads		
Dependent variable	Actual caseloads	Fixed caseloads	Judicial elections	Mid-term leaves	County rotations
Panel A. Variance decomposition for change in county stringency (ΔS_{st})					
SD of change in county stringency (ΔS_{ct})	0.0180	0.0140	0.0077	0.0064	0.0032
Variance of ΔS_{ct}	32.5E-5	19.6E-5	5.9E-5	4.1E-5	1.0E-5
% of total variance in ΔS_{ct}	100%	60%	18%	13%	3%
Panel B. Main effects (scaled to 1 SD of county stringency in each measure)					
Δ log # active sentences	0.208^{***}	0.166^{**}	0.102	0.270^{***}	0.207
	(0.067)	(0.064)	(0.071)	(0.102)	(0.172)
Δ mean Math score	-0.024^{***}	-0.036^{**}	-0.035^{*}	-0.100^{**}	0.042
	(0.008)	(0.018)	(0.018)	(0.044)	(0.052)
Δ mean English score	-0.014^{**}	-0.020^{*}	-0.027^{**}	-0.046^{**}	0.033
	(0.007)	(0.010)	(0.013)	(0.023)	(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
Panel C. Balance tests (F statistics)					
Criminal case totals	0.3	1.5	2.3	0.5	1.2
Defendant characteristics	1.0	1.2	1.6	2.2	2.2
Student characteristics	1.5	1.0	0.6	0.7	1.5
Prior year test scores	0.5	2.0	0.9	1.0	1.3

Notes: This table decomposes the change in county stringency, ΔS_{ct}, into its underlying components, and estimates the effects of each component on student test scores. We begin in column (A) with the actual caseload measure, which uses variation in both whether or not a judge serves in a county in a given year as well as their number of cases that county/year. Column (B) reports the fixed caseload measure, which restricts to only the variation in whether or not a judge serves in a county (by using, in each year, the judge's mean caseload across all years). Columns (C)–(E) decompose the fixed caseload measure in column (B) into three components. Column (C) computes ΔS_{ct} using only judges who began or stopped working in *any* county in our data due to an election win or loss. We compute this measure by linking judges in the court data to biannual election results using judge initials, names, and districts. Column (D) uses variation from mid-term leaves, which we define as all reasons that judges began or stopped working in *any* county in our data other than elections. 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, ΔS_{ct} (first row). This panel also reports the percent of the total variance in ΔS_{ct} explained by each component, which is equal to the variance of ΔS_{ct} in each column divided by the variance of ΔS_{ct} in column (A). Panel B shows the impacts of each of these underlying forms of variation on active sentences and test scores; these are analogous to the coefficients in Tables 3–4, but we use ΔS_{ct} computed from the source of variation highlighted in each column. We normalize each ΔS_{ct} measure so that one unit corresponds to one standard deviation in county stringency for that measure. Parentheses contain standard errors clustered at the county level. Panel C reports F-statistics from balance tests for the joint significance of changes in criminal case totals, defendant characteristics, student characteristics, and prior year test scores. These balance tests are analogous to those in Tables 3 and 5, but again use ΔS_{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

	(A)	(B)	(C)	(D)	(E)
				Judge stringency estimate	
Dependent variable	Mean	OLS coef. on active sentence	Event study estimate	Reduced form (SD units)	IV for active sentence
Panel A. First stage					
Active sentence	0.137			0.034^{***}	
				(0.003)	
Panel B. Academic performance					
Math score	−0.359	-0.034^{***}	-0.015^{***}	−0.006	−0.164
		(0.010)	(0.005)	(0.006)	(0.169)
English score	−0.308	−0.014	-0.012^{**}	-0.012^{**}	-0.348^{**}
		(0.011)	(0.005)	(0.006)	(0.164)
Panel C. Attendance and misbehavior					
Days absent	9.312	0.412^{***}	0.456^{***}	0.079^{*}	2.372^{*}
		(0.094)	(0.059)	(0.045)	(1.396)
Any suspension	0.158	0.009^{***}	0.013^{***}	0.003^{**}	0.099^{**}
		(0.003)	(0.002)	(0.001)	(0.043)
Number of suspension days	1.675	0.153^{***}	0.252^{***}	0.021	0.642
		(0.055)	(0.036)	(0.022)	(0.666)
Fighting incident	0.048	0.003^{**}	0.002^{*}	0.001^{*}	0.042^{*}
		(0.002)	(0.001)	(0.001)	(0.024)
Panel D. Neighborhood quality and student criminal activity					
Neighborhood SES percentile	0.454	-0.024^{***}	0.003^{*}	0.003	0.086
		(0.004)	(0.002)	(0.002)	(0.066)
Student has criminal charge	0.004	0.002	−0.002	−0.000	−0.012
		(0.002)	(0.001)	(0.001)	(0.016)
Judge stringency sample	Y	Y		Y	Y
Event study sample			Y		
Years, t, relative to case disposition, $\tau(i)$	$t \geq \tau(i)$	$t \geq \tau(i)$	All t	$t \geq \tau(i)$	$t \geq \tau(i)$
N (Math scores)	296,242	296,242	699,084	296,242	296,242
# students	118,416	118,416	128,829	118,416	118,416
F statistics:					
First stage				103.0	
Defendant balance test				0.9	
Student balance test				1.1	

Notes: This table presents estimates of the direct impacts of a household incarceration using our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our direct judge stringency sample (column (B) in Appendix Table A3). Column (C) includes students in our direct event study sample (column (A) in Appendix Table A3). See Appendix C.5 for details on our samples. Each row corresponds to a separate regression using the dependent variable listed in the first column. Regressions are at the student × year level.

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

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

TABLE 8: Indirect impacts of household incarceration on classmates

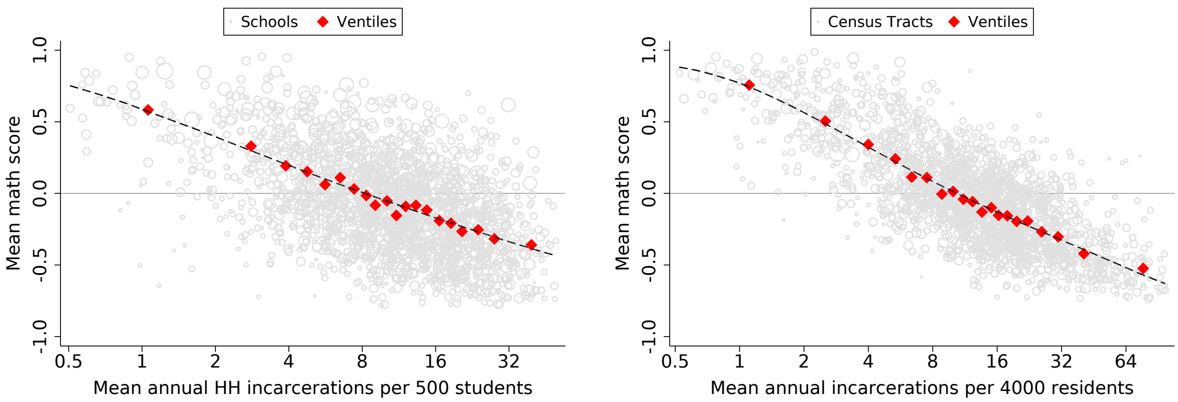
	(A)	(B)	(C)	(D)	(E)
				Judge stringency estimate	
Dependent variable	Mean	OLS coef. on active sentence	Event study estimate	Reduced form (SD units)	IV for active sentence
Panel A. First stage					
Active sentence	0.133			0.031^{***} (0.002)	
Panel B. Academic performance					
Math score	−0.117	-0.006^{*} (0.003)	-0.004^{**} (0.002)	−0.001 (0.002)	−0.017 (0.062)
English score	−0.048	-0.008^{***} (0.003)	-0.003^{**} (0.001)	−0.001 (0.002)	−0.047 (0.053)
Judge stringency sample	Y	Y		Y	Y
Event study sample			Y		
Years, t, relative to case disposition, $\tau(i)$	$t \geq \tau(i)$	$t \geq \tau(i)$	All t	$t \geq \tau(i)$	$t \geq \tau(i)$
N (Math scores)	32,120,520	32,120,520	103,609,169	32,120,520	32,120,520
# students	1,454,577	1,454,577	1,470,600	1,454,577	1,454,577
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 our event study and judge stringency strategies. Columns (A)–(B) and (D)–(E) include students in our indirect judge stringency sample (column (E) in Appendix Table A3). Column (C) includes students in our indirect event study sample (column (D) in Appendix Table A3). See Appendix C.5 for details on our samples. Each row corresponds to a separate regression using the dependent variable listed in the first column. Regressions are at the student × directly-impacted child × year level; thus student × year observations are repeated for each directly-impact child in their school/grade.

Column (A) displays means of each outcome variable measured in calendar years, t, in or after the disposition year of the defendant's case, $\tau(i)$. Column (B) shows OLS estimates of β from Equation 7 using outcomes measured in years $t \geq \tau(i)$. Column (C) displays estimates of δ from the event study specification 6 using all years t. Column (D) shows reduced-form judge stringency estimates of γ from Equation 8 for years $t \geq \tau(i)$, with estimates normalized to represent a one standard deviation increase in judge stringency. Column (E) shows estimates of β from the IV specification based on equations 7 and 8 using years $t \geq \tau(i)$. Regressions in columns (B), (D), and (E) include court × year × 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 B1 for details on the balance tests.

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

Figures

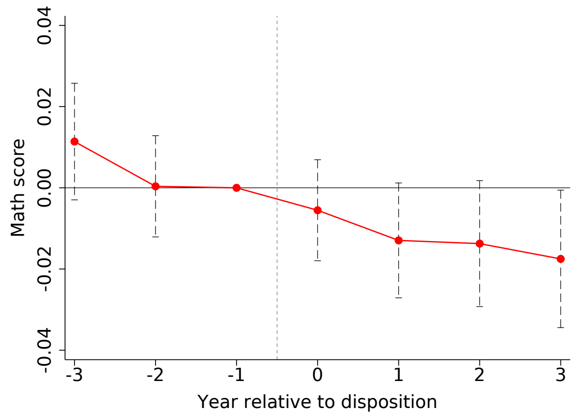


Panel A. School exposure

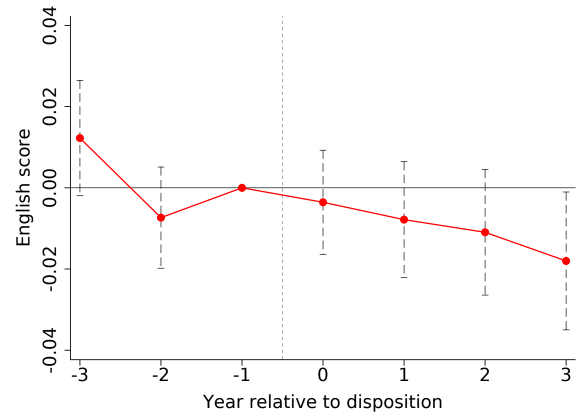
Panel B. Neighborhood exposure

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

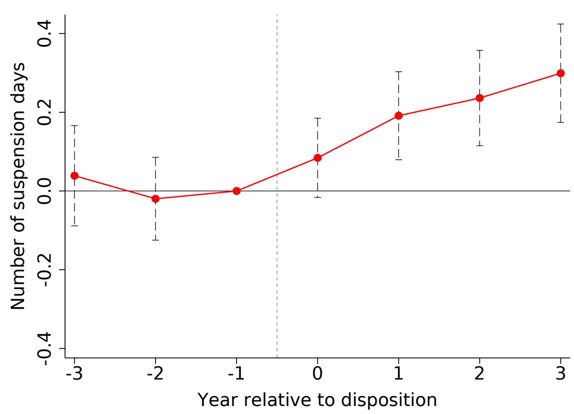
Notes: This figure plots the relationship between student math scores and exposure to incarceration at the school (Panel A) and neighborhood (Panel B) levels. In Panel A, grey circles represent schools, and the x-axis shows the average annual number of active sentences for defendants linked to students 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 99th and 1st percentiles, and use log scales on the x axis excluding observations below 0.5 mean annual incarcerations per 500 students/4,000 residents.



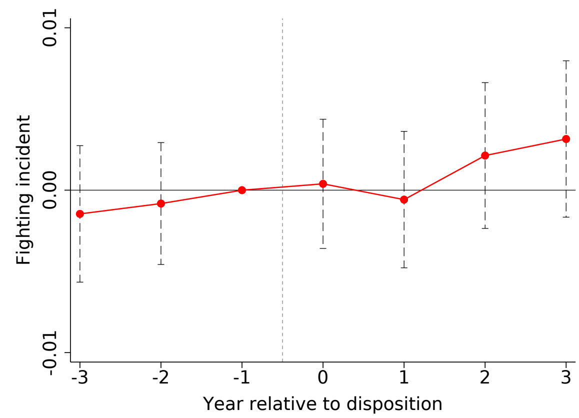
Panel A. Math score



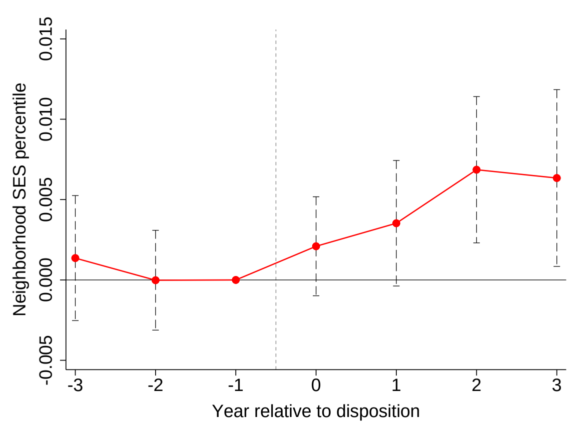
Panel B. English score



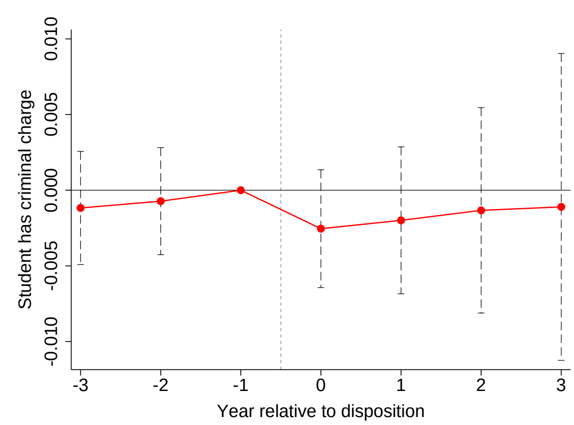
Panel C. Number of suspension days



Panel D. Fighting incident



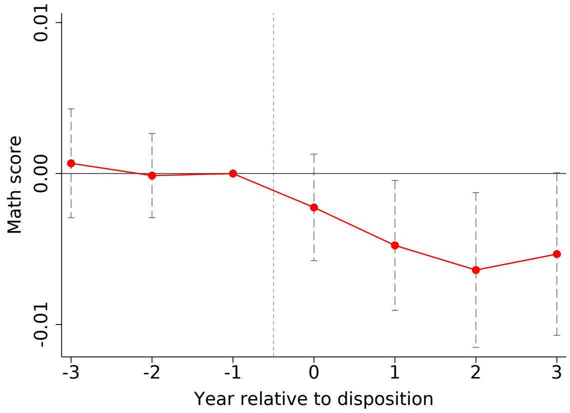
Panel E. Neighborhood SES percentile



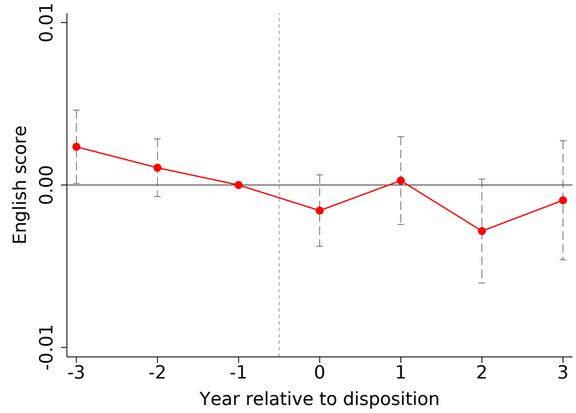
Panel F. Student has criminal charge

FIGURE 2: Direct impacts of household incarceration — Event study

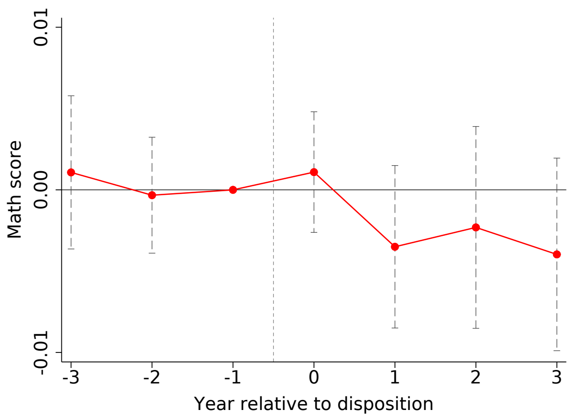
Notes: This figure shows event studies of the direct impacts of household incarceration on student outcomes. We use our direct exposure event study sample (column (A) in Appendix Table A3) and estimate the following regression: $y_{it} = \sum_{l=-8}^{9} \delta_l \mathbb{1}\{t - \tau(i) = l\} \times \text{Active Sentence}_{k(i)} + \eta_i + \lambda_{og\tau(i)t} + \varepsilon_{it}$. 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 $\tau(i)$. Active Sentence$_{k(i)}$ is an indicator equal to one if defendant $k(i)$ received an active sentence. We interact this term with dummies for years l relative to the defendant's disposition, omitting the $l = -1$ term. η_i is an individual fixed effect. $\lambda_{og\tau(i)t}$ is an offense class $\times$ academic cohort $\times$ disposition year $\times$ calendar year fixed effect. The graphs plot the δ_l coefficients from $l = -3$ to 3. The outcome variable for each regression, y_{it}, is listed in the panel title; see Appendix C.1 for details on variable definitions. Dashed lines are 95 percent confidence intervals using standard errors clustered at the defendant level.



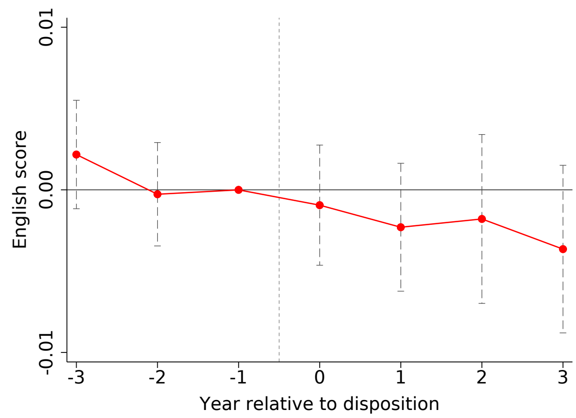
Panel A. Math score (Event study)



Panel B. English score (Event study)



Panel C. Math score (Judge stringency)



Panel D. English score (Judge stringency)

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 6: $y_{ikt} = \sum_{l=-8}^{9} \delta_l \cdot \mathbb{1}\{t - \tau(k) = l\} \times \text{Active Sentence}_{d(k)} + \eta_i + \lambda_{og\tau(k)t} + \varepsilon_{ikt}$. 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 $\tau(k)$. Active Sentence$_{d(k)}$ is an indicator equal to one if defendant $d(k)$ received an active sentence. η_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 δ_l coefficients, where the x-axis denotes years l relative to the defendant's disposition. We omit the $l = -1$ interaction term, so δ_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 8, and interact the stringency measure with dummies for years l relative to the defendant's disposition. The graphs plot the γ_l coefficients with years since disposition, l, on the x-axis. We omit the $l = -1$ interaction term, so γ_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; see Appendix C.1 for details on variable definitions. 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.

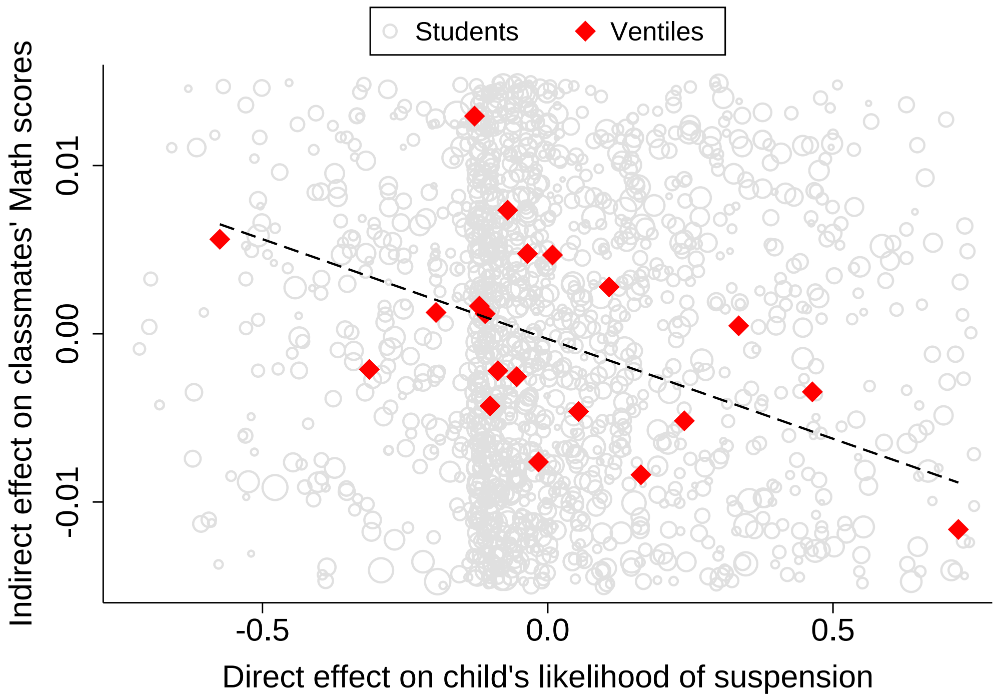


FIGURE 4: Classroom disruption channel of indirect academic impacts

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.

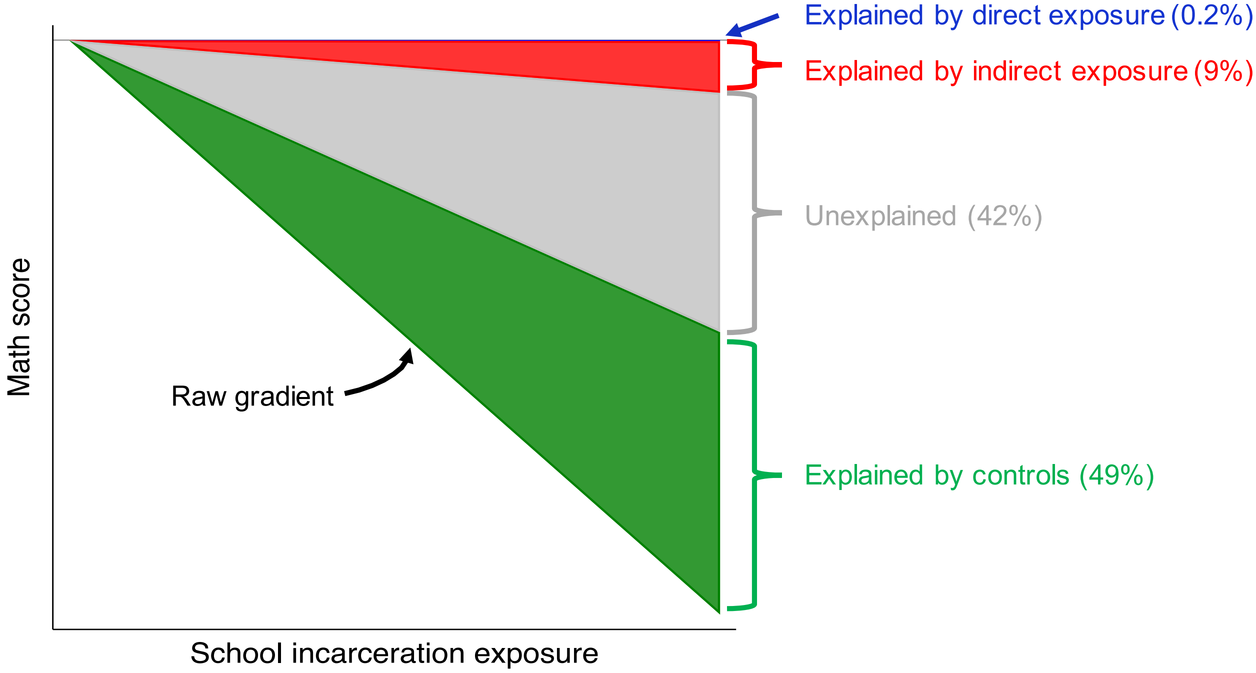


FIGURE 5: Decomposition of the gradient between math scores and school incarceration exposure

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 A7 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 Δ county stringency (SD units)				
Panel A. Heterogeneity by school exposure to criminal cases					
		Quartiles of criminal cases per student in 2010			
Dependent variable	All schools	Bottom Qtile	Q2	Q3	Top Qtile
Δ log HH incarcerations per student	0.126***	0.121	0.139*	0.075	0.296**
	(0.038)	(0.141)	(0.079)	(0.051)	(0.109)
Δ HH incarcerations per 500 students	1.031**	0.562	1.481	0.187	4.657***
	(0.485)	(1.317)	(1.124)	(0.456)	(1.475)
Δ mean Math score	−0.028***	−0.025	−0.015	−0.027**	−0.059**
	(0.008)	(0.048)	(0.009)	(0.009)	(0.024)
Δ mean English score	−0.016**	−0.005	−0.022**	−0.005	−0.028
	(0.006)	(0.038)	(0.010)	(0.004)	(0.025)
N (# school/years)	9,112	2,324	2,236	2,288	2,264
Panel B. Heterogeneity by neighborhood exposure to criminal cases					
		Quartiles of criminal cases per resident in 2010			
Dependent variable	All tracts	Bottom Qtile	Q2	Q3	Top Qtile
Δ log incarcerations per resident	0.142***	0.172**	0.132***	0.138***	0.095**
	(0.031)	(0.070)	(0.033)	(0.036)	(0.040)
Δ incarcerations per 4000 residents	2.332***	1.254***	1.897***	2.321***	3.622**
	(0.546)	(0.279)	(0.405)	(0.689)	(1.742)
Δ mean Math score	−0.026***	−0.019**	−0.026**	−0.024**	−0.028***
	(0.008)	(0.009)	(0.011)	(0.011)	(0.009)
Δ mean English score	−0.016***	−0.020**	−0.012*	−0.016	−0.018*
	(0.006)	(0.010)	(0.007)	(0.011)	(0.010)
N (# tract/years)	8,596	2,152	2,148	2,148	2,148

Notes: This table examines heterogeneity in the effects of county stringency on student test scores by exposure to criminal cases. The table displays estimates of β from Equation 2, but outcomes are defined at the school (Panel A) or Census tract (Panel B) level rather than 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: County stringency effects under different sources of variation in judge turnover

	(A)	(B)	(C)	(D)	(E)
		Sources of variation in fixed caseloads			
Dependent variable	Fixed caseloads	Close elections	Elections due to retiring judges	Election defeats	Using retiring judges only
Panel A. Variance decomposition for change in county stringency (ΔS_{st})					
SD of change in county stringency (ΔS_{ct})	0.0140	0.0068	0.0048	0.0055	0.0048
Variance of ΔS_{ct}	19.6E-5	4.7E-5	2.3E-5	3.0E-5	2.3E-5
% of total variance in ΔS_{ct}	100%	24%	12%	16%	12%
Panel B. Main effects (scaled to 1 SD of county stringency in each measure)					
Δ log # active sentences	0.166^{**}	0.173^{**}	0.233^{*}	0.034	0.160
	(0.064)	(0.082)	(0.129)	(0.081)	(0.115)
Δ mean Math score	-0.036^{**}	-0.043^{*}	-0.028	-0.050^{**}	-0.042
	(0.018)	(0.024)	(0.035)	(0.021)	(0.039)
Δ mean English score	-0.020^{*}	-0.025	-0.042	-0.022	-0.024
	(0.010)	(0.015)	(0.030)	(0.015)	(0.023)
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

Notes: This table decomposes the change in county stringency, ΔS_{ct}, into its underlying components, and estimates the effects of each component on student test scores. We begin in column (A) with our fixed caseload measure, which uses 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 (B)–(E) decompose the fixed caseload measure in column (A) into four components. Column (B) computes ΔS_{ct} using only turnover from close elections, defined as elections in the winning judge's share was below 55 percent. Column (C) computes ΔS_{ct} using only election turnover that occurred when the incumbent judge retired. Column (D) computes ΔS_{ct} using only election turnover that occurred when the incumbent judge lost the election. Column (E) uses only the retiring judge's stringency to compute ΔS_{ct}, i.e., we do not use the stringency of the judge that replaced the retiring judge. We identify elections and the reasons for election turnover by linking judges in the court data to biannual election results using judge initials, names, and districts. For column (E), we define retirements to also include judges who left the court data in the middle of their term.

Panel A reports the standard deviation of the *change* in county stringency, ΔS_{ct} (first row). This panel also reports the percent of the total variance in the fixed caseload measure of ΔS_{ct} that is explained by each component, which is equal to the variance of ΔS_{ct} in each column divided by the variance of ΔS_{ct} in column (A). Panel B shows the impacts of each of these underlying forms of variation on active sentences and test scores; these are analogous to the coefficients in Tables 3–4, but we use ΔS_{ct} computed from the source of variation highlighted in each column. We normalize each ΔS_{ct} measure so that one unit corresponds to one standard deviation in county stringency for that measure. Parentheses contain standard errors clustered at the county level.

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

TABLE A3: Summary statistics for direct and indirect analysis samples

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

Notes: This table displays summary statistics on our direct and indirect analysis samples for Sections 4–5. We show student characteristics (Panel A), student outcomes (Panel B), and characteristics of matched defendants and their offenses (Panel C). Column (A) shows our event study sample for the direct impacts of household incarceration, which includes students whose linked defendant was convicted of their most serious offense. Column (B) shows our judge stringency sample for direct impacts, which includes students whose linked defendant faced a judge for which we can compute a stringency measure. Column (C) displays mean complier characteristics for the judge stringency sample in column (B) using the methodology of Frandsen et al. (2019); we estimate our judge stringency IV specification (Equations 3 and 5) where the dependent variable is the student/defendant characteristic interacted with an indicator for an active sentence, and display the resulting IV coefficient in column (C). Columns (D)–(E) include students who were in the same school and grade as the children in columns (A)–(B) in the year of the defendant's disposition; these are our samples for examining the indirect impacts of 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 averaged over all years in 2006–2017. The defendant/offense characteristics in Panel C correspond to the most serious offense linked to each child, as described in Appendix C.5. See Appendix C.1 for details on variable definitions.

TABLE A4: Event study robustness tests

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

Notes: This table examines robustness for our event study estimates of the direct and indirect effects of exposure to household incarceration. Panel A presents estimates of δ from the direct event study regression 4. Panel B presents estimates of δ from the indirect event study regression 6.

Column (C) presents estimates from our benchmark specification, which replicates the results in column (C) of Tables 7 and 8. Our benchmark specification includes offense class at conviction (o) $\times$ academic cohort (g) $\times$ disposition year (τ) $\times$ calendar year (t) fixed effects, which we denote by $\lambda_{og\tau(i)t}$. The other columns of this table present results using different definitions of this fixed effect term. Column (A) excludes interactions for offenses, o, and thus the fixed effects are $\lambda_{g\tau(i)t}$. Column (B) defines offenses, o, based on the offense class at the time of arrest (rather than at conviction). Column (D) defines offenses, o, by the 4-digit offense code at conviction rather than the offense class at conviction. Column (E) is identical to column (D), except we add a covariate for the defendant's prior points at conviction interacted with a linear term for years since disposition, $t - \tau(i)$.

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

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

TABLE A5: Judge stringency robustness tests

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

Notes: This table examines robustness for our judge stringency estimates of the direct and indirect effects of exposure to household incarceration. Panel A presents reduced form estimates of γ from the direct judge stringency regression 5. Panel B presents reduced form estimates of γ from the indirect judge stringency regression 8. Column (A) presents estimates from our benchmark specification, which replicates the results in column (D) of Tables 7 and 8. For column (B), we compute stringency at the judge × 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 × year × offense class (rather than dummies for the structured sentencing grid) in our leave-out sample, and then average the residuals from this regression at the judge × year level. Column (D) is identical to column (A), except that in the fixed effect term, $\theta_{oc\tau}$, 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 A6: Heterogeneity in indirect event study estimates by direct effects on child behavior

	(A)	(B)	(C)
	Behavioral outcome for directly-impacted child		
Independent variables	Any suspension	Number of suspension days	Fighting incident
Panel A. Dependent variable: Math scores of indirectly-impacted children			
Active sentence × Post disposition	-0.0040^{**}	-0.0039^{**}	-0.0041^{**}
	(0.0019)	(0.0019)	(0.0019)
Active × Post × Direct effect on any suspension	-0.0112^{***}		
	(0.0042)		
Active × Post × Direct effect on # suspension days		-0.0005^{***}	
		(0.0002)	
Active × Post × Direct effect on fighting incident			-0.0112^{*}
			(0.0065)
N (Math scores)	102,146,692	102,146,692	102,146,692
Panel B. Dependent variable: English scores of indirectly-impacted children			
Active sentence × Post disposition	-0.0025^{**}	-0.0026^{**}	-0.0026^{**}
	(0.0013)	(0.0013)	(0.0013)
Active × Post × Direct effect on any suspension	-0.0052^{*}		
	(0.0029)		
Active × Post × Direct effect on # suspension days		-0.0001	
		(0.0001)	
Active × Post × 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 6, except we include the triple interaction term $\mathbb{1}\{t \geq \tau(i)\} \times$ Active Sentence$_i \times \hat{\delta}_i$. This table displays the coefficient on both the double interaction, $\mathbb{1}\{t \geq \tau(i)\} \times$ 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 A7: Decomposition of the gradient between test scores and school incarceration exposure

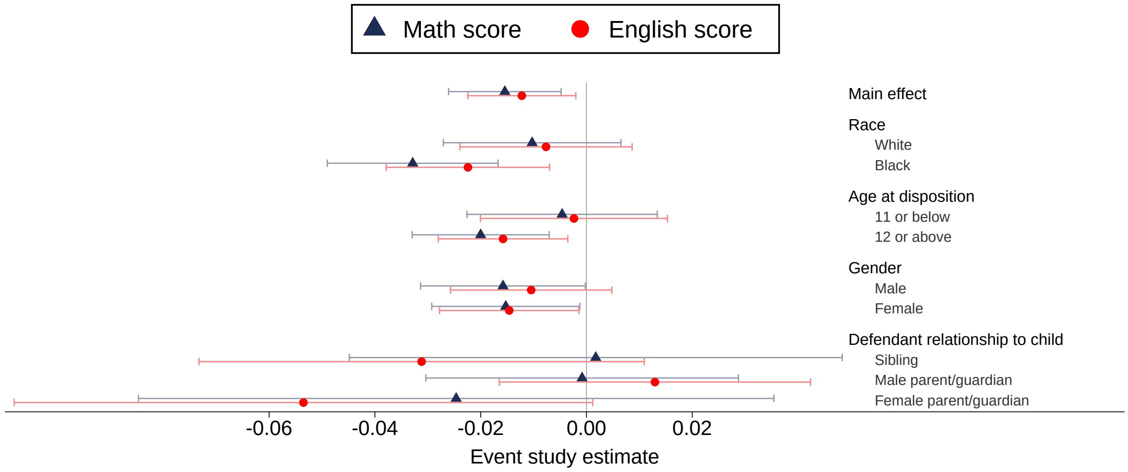
	(A)	(B)	(C)	(D)
Panel A. Students and incarceration exposure				
	Individual	School/grade	School	
Mean # students	1.0	192.2	715.2	
Mean # HH incarcerations/year	0.023	4.5	16.7	
Panel B. Estimates from paper				
	Effect of 1 HH incarceration in event study		Effect of 1 log point in HH incarcerations/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

Notes: This table uses event study estimates from the paper to decompose the gradient between test scores and school incarceration exposure into direct, indirect, and residual channels.

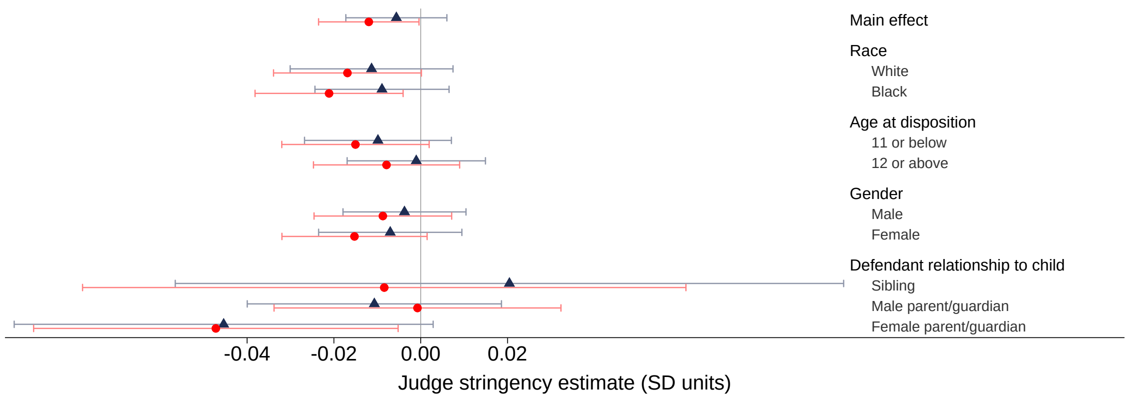
Panel A reports summary statistics at the individual (column (A)), school/grade (column (B)), and school (column (C)) levels. We report the mean number of students in each group across all North Carolina public school students in our data from 2010–2014. We also report the mean number of household incarcerations in each group. For example, the average public school student experienced 0.023 household incarcerations in a typical year, but were in the same school/grade as 4.5 students with a household incarceration.

Panel B summarizes estimates from the paper. Column (A) displays event study estimates of the direct impacts of 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 (C) 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).



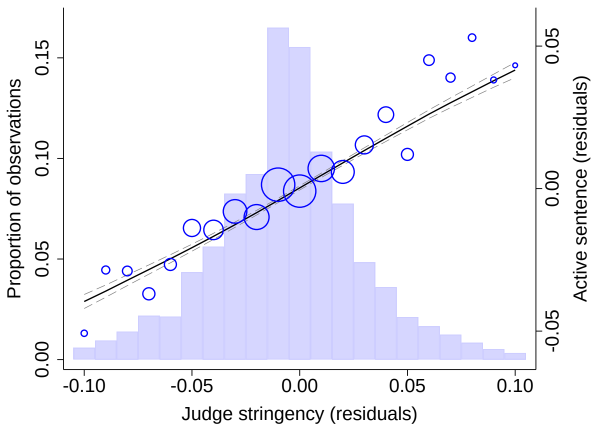
Panel A. Event study



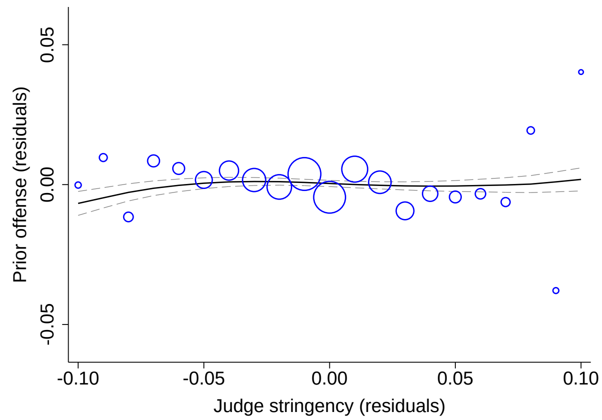
Panel B. Judge stringency

FIGURE A1: Heterogeneity in direct effects of household incarceration on test scores

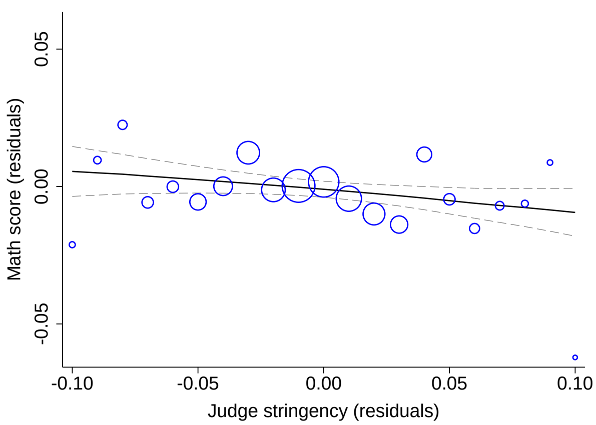
Notes: This figure displays heterogeneity in the direct impacts of household incarceration on student test scores. Panel A presents event study estimates using our direct exposure sample from column (D) in Table 1. Panel B presents reduced-form judge stringency estimates using our direct exposure sample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



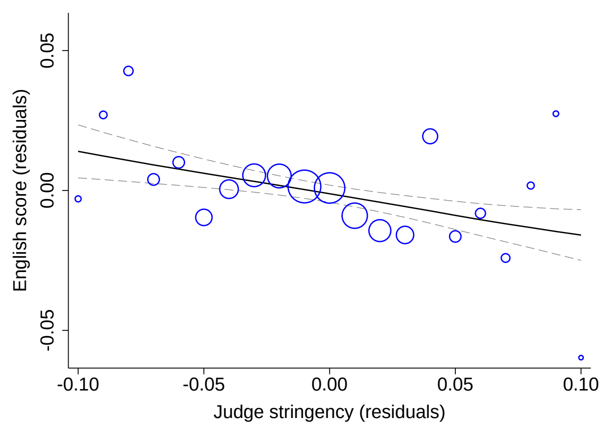
Panel A. Active sentence



Panel B. Prior offense



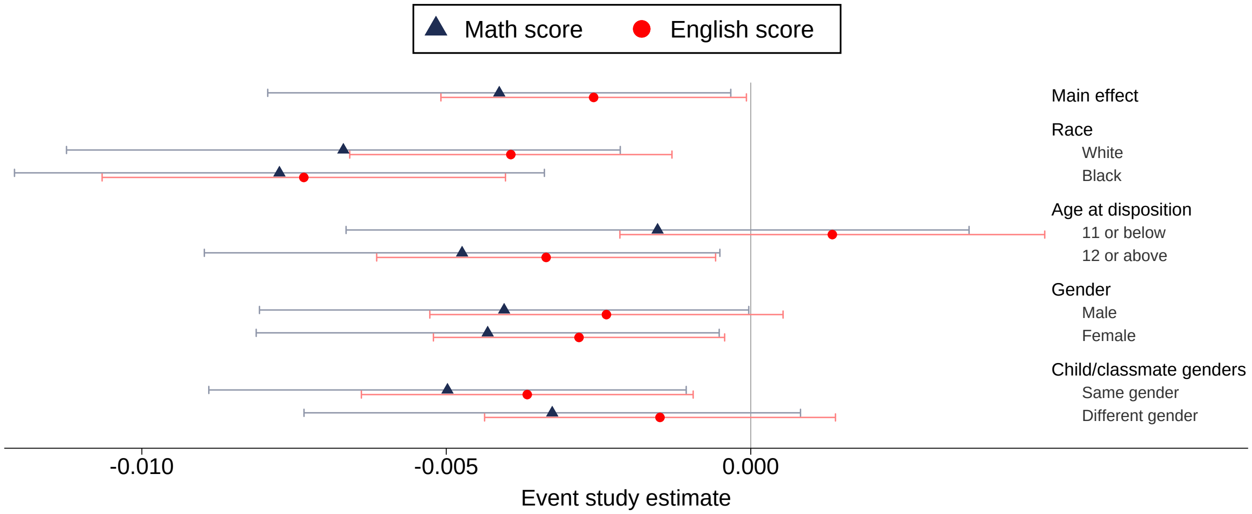
Panel C. Math score (directly-impacted child)



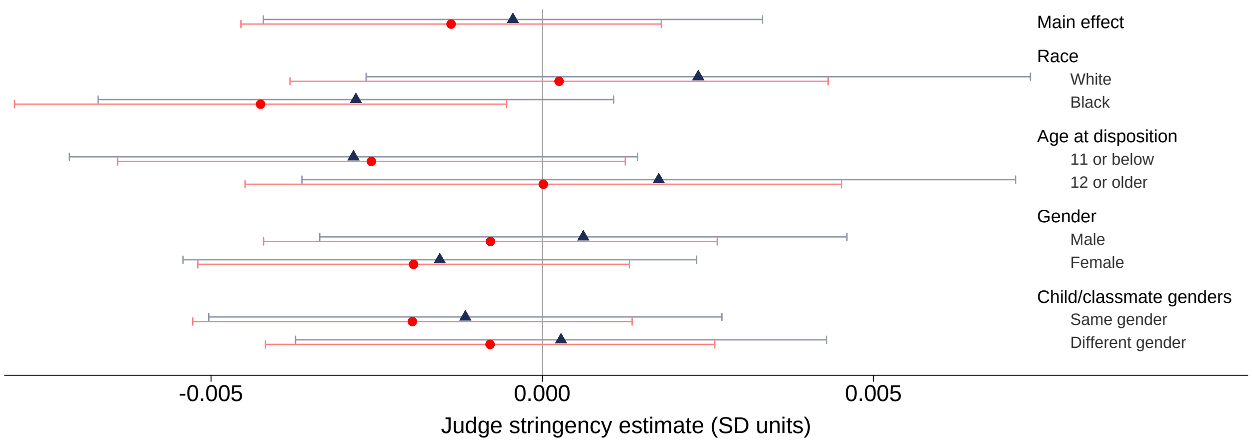
Panel D. English score (directly-impacted child)

FIGURE A2: Reduced form judge stringency effects

Notes: This figure shows the distribution of judge stringency and its relationship with defendant and student outcomes. Each panel depicts the relationship between outcome residuals (y-axis) and judge stringency residuals (x-axis). We compute these residuals from a regression of each variable on court $\times$ year $\times$ offense class dummies. The sample includes students in our direct judge stringency sample (column 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).



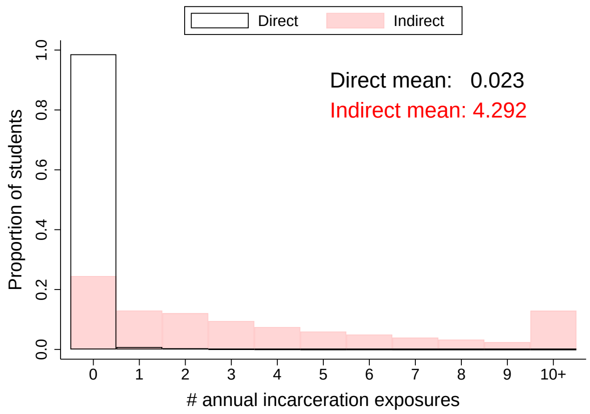
Panel A. Event study



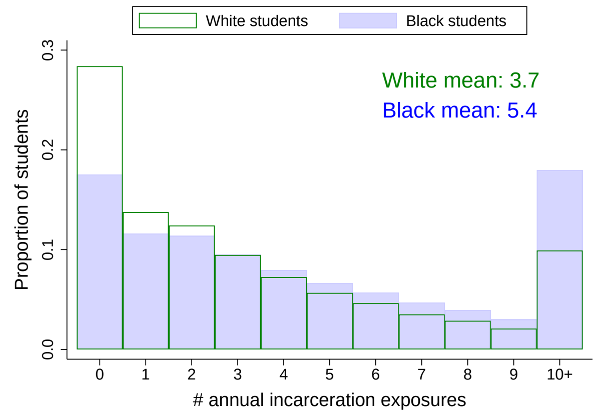
Panel B. Judge stringency

FIGURE A3: 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.



Panel A. Direct and indirect exposure



Panel B. Indirect exposure by race

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

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

B The Role of Judges in the North Carolina Court System

B.1 Court Structure and Judge Turnover

North Carolina's criminal court system is organized into two divisions: the Appellate Division, and the Trial Division comprising the superior and division courts.[29] We focus on district and superior courts, which hear new criminal cases. Superior courts have jurisdiction over all felony cases that are heard by trial, as well as misdemeanor cases that are appealed from district court. District courts have jurisdiction over all new misdemeanor cases. Felony cases often begin in District Court for pre-trial proceedings; class H or class I felonies may also be resolved in district court with a guilty plea. Superior court trials are presided over by a judge with a jury of twelve, whereas district court trials are always held without jury.

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

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

[29] For details on the North Carolina court structure, see: https://www.nccourts.gov/courts.

B.2 Timeline of Court Process

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

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

Since the 1994 Structured Sentencing Act, judicial sentencing has been limited under a grid system designed to limit the extent of judicial discretion, as in Figure B2. These establish minimum and maximum ranges for sentencing based on offense severity and the number of prior convictions. They also establish the different sentencing options available to judges: 1) community punishment (probation), 2) intermediate punishment (probation with additional conditions[30]), and 3) active punishment—incarceration in prison or jail. An active sentence necessarily implies judicial incarceration for at least the minimum sentence length.[31] 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.

[30] 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.

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

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

B.3 Assignment of Cases to Judges

The assignment of cases to judges is determined by county clerks. At the district court level, the senior judge, in conjunction with the county clerk, determines an assignment of judges to specific courtrooms specializing in different categories of cases.[32] 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.[33]

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

[32] 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.

[33] 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%20Of%20Term_Out%20Of%20Session_Out%20of%20County.pdf.

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

MISDEMEANOR PUNISHMENT CHART

<table>
<tr><th rowspan="3">CLASS</th><th colspan="4">PRIOR CONVICTION LEVEL</th></tr>
<tr><th>I</th><th colspan="2">II</th><th>III</th></tr>
<tr><th>No Prior Convictions</th><th colspan="2">One to Four Prior Convictions</th><th>Five or More Prior Convictions</th></tr>
<tr><td>A1</td><td>C/I/A
1 - 60 days</td><td colspan="2">C/I/A
1 - 75 days</td><td>C/I/A
1 - 150 days</td></tr>
<tr><td>1</td><td>C
1 - 45 days</td><td colspan="2">C/I/A
1 - 45 days</td><td>C/I/A
1 - 120 days</td></tr>
<tr><td>2</td><td>C
1 - 30 days</td><td colspan="2">C/I
1 - 45 days</td><td>C/I/A
1 - 60 days</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">C
Fine Only*
1 - 10 days</td><th>One to Three Prior Convictions</th><th>Four Prior Convictions</th><td rowspan="2">C/I/A
1 - 20 days</td></tr>
<tr><td>C
Fine Only*
1 - 15 days</td><td>C/I
1 - 15 days</td></tr>
</table>

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

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

FIGURE B1: Range of Outcomes under North Carolina 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 *****

FELONY PUNISHMENT CHART

PRIOR RECORD LEVEL

OFFENSE CLASS

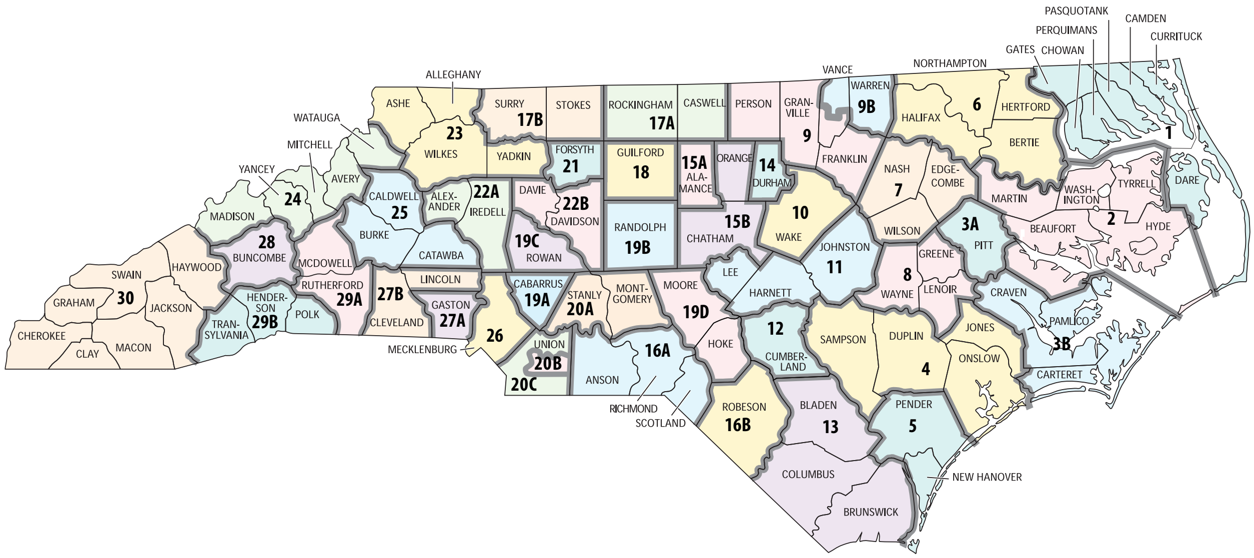
	I 0-1 Pt	II 2-5 Pts	III 6-9 Pts	IV 10-13 Pts	V 14-17 Pts	VI 18+ Pts
A	Death or Life Without Parole Defendant Under 18 at Time of Offense: Life With or Without Parole					
B1	A	A	A	A	A	A
	240 - 300	*276 - 345*	*317 -397*	*365 - 456*	*Life Without Parole*	*Life Without Parole*
	192 - 240	**221 - 276**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

DISPOSITION
Aggravated Range
PRESUMPTIVE RANGE
Mitigated Range

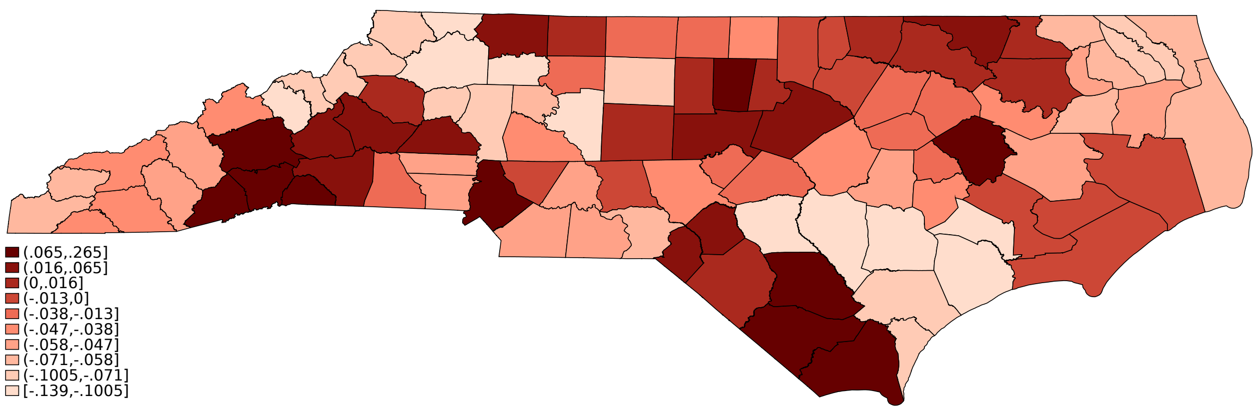
A – Active Punishment **I – Intermediate Punishment** **C – Community Punishment**
Numbers shown are in months and represent the range of minimum sentences

FIGURE B2: Range of Outcomes under North Carolina 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.



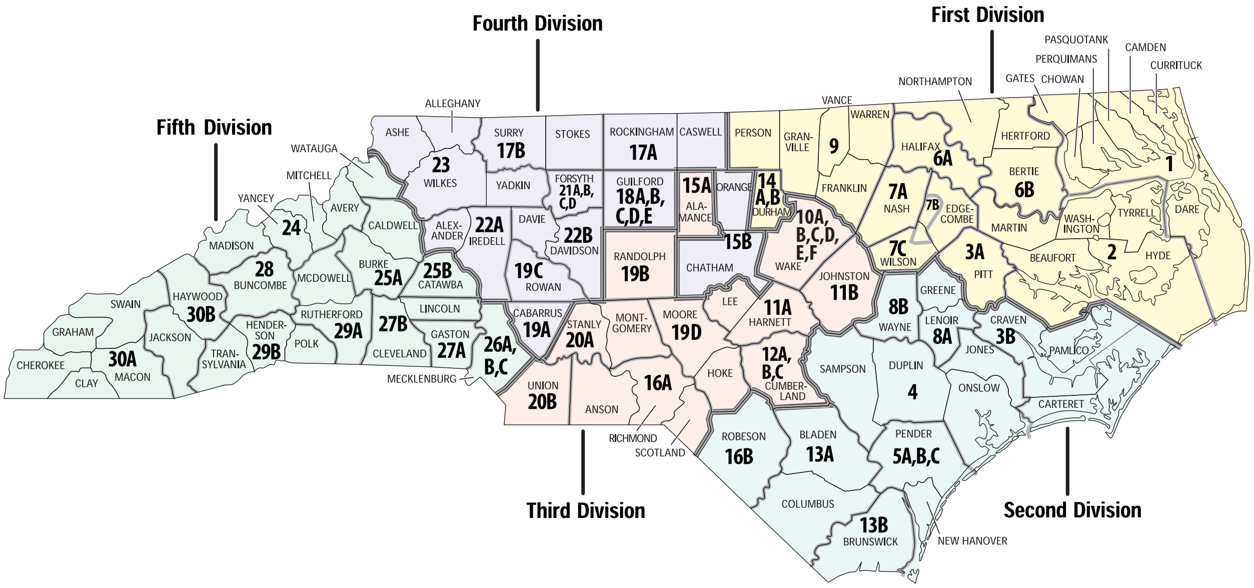
Panel A. District court districts (Effective January 1, 2019)



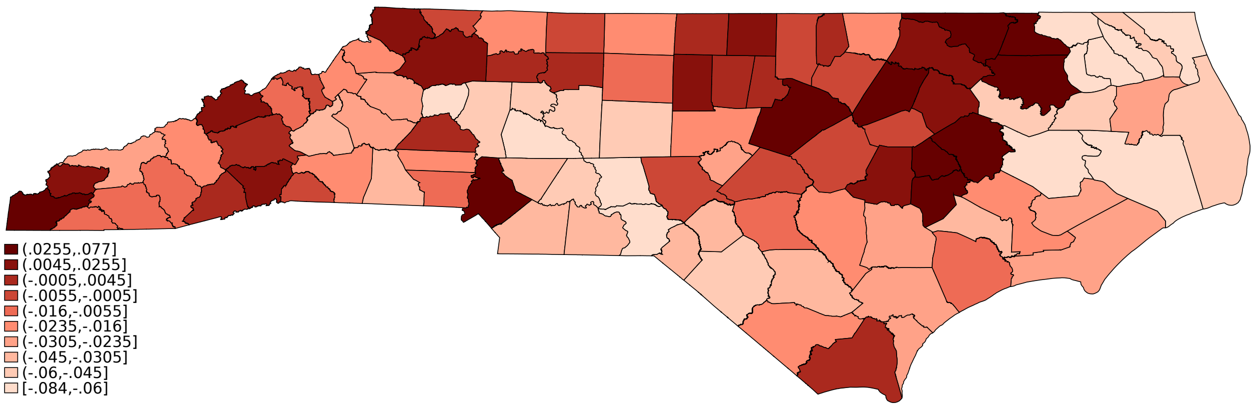
Panel B. Mean stringency for district court judges (2010)

FIGURE B3: North Carolina District Courts

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



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



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

FIGURE B4: North Carolina Superior Courts

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

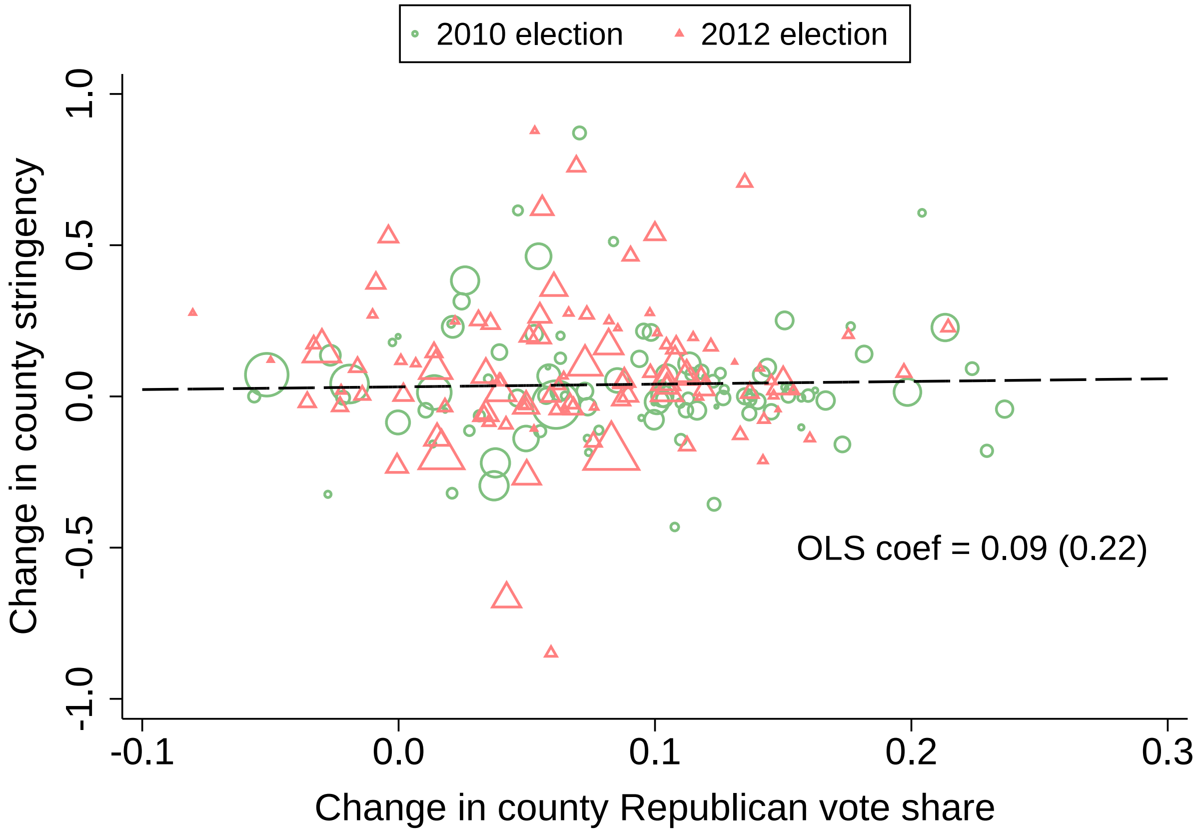
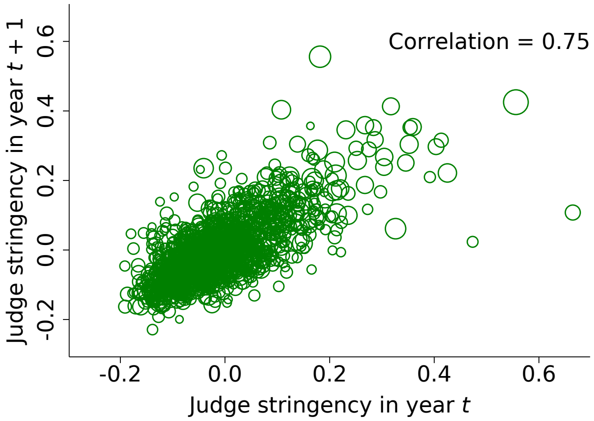
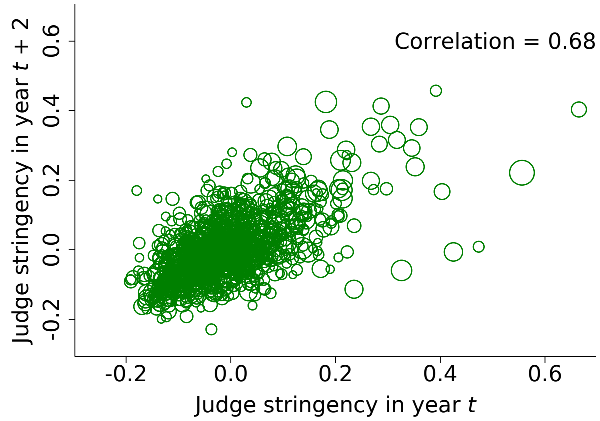


FIGURE B5: Changes in county stringency and county-level Republican vote share

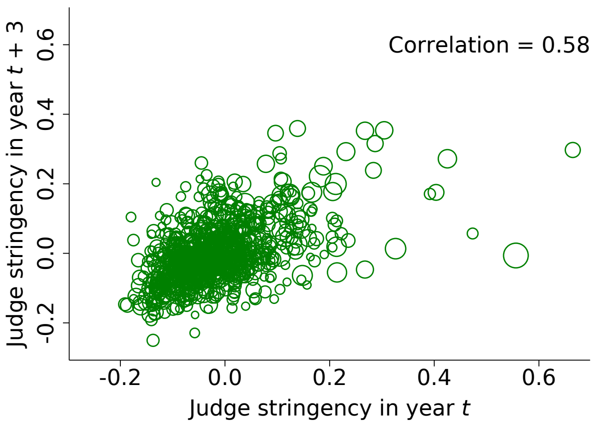
Notes: This figure plots the relationship between changes in county stringency and county-level Republican vote share. The x-axis shows the change in the Republican vote share between years t and $t-4$, where t includes the November 2010 election (green circles) and the November 2012 election (red triangles).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



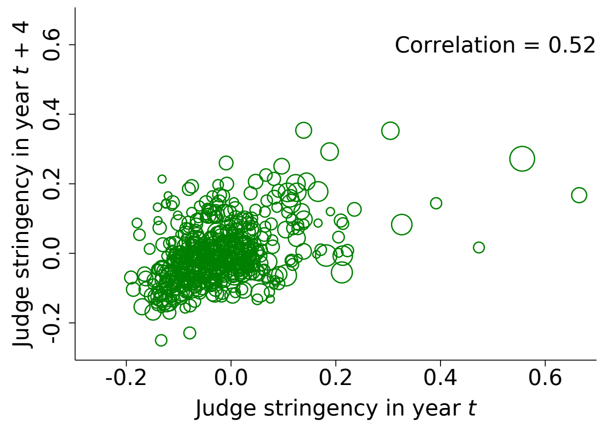
Panel A. Years t and $t + 1$



Panel B. Years t and $t + 2$



Panel C. Years t and $t + 3$



Panel D. Years t and $t + 4$

FIGURE B6: Persistence of individual judge stringency

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

TABLE B1: Judge stringency balance tests

	(A)	(B)	(C)	(D)
	Dependent variable in direct sample		Dependent variable in indirect sample	
Covariate	Active sentence	Judge stringency (SD units)	Active sentence	Judge stringency (SD units)
Panel A. Defendant/offense characteristics				
Male	0.040***	0.006	0.038***	0.008
	(0.003)	(0.005)	(0.002)	(0.006)
Age at disposition	0.001***	−0.000	0.001***	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
White	−0.031***	−0.002	−0.028***	−0.002
	(0.005)	(0.009)	(0.004)	(0.008)
Black	−0.005	0.000	−0.002	0.002
	(0.005)	(0.011)	(0.004)	(0.009)
Multiple offenses	0.026***	0.009	0.026***	0.015**
	(0.003)	(0.006)	(0.003)	(0.006)
Prior offense	0.014***	−0.002	0.015***	−0.002
	(0.003)	(0.005)	(0.003)	(0.006)
Prior active sentence	0.231***	0.019**	0.235***	0.017
	(0.010)	(0.009)	(0.010)	(0.012)
F statistic: All coefficients zero	97.8	1.1	136.9	1.3
Panel B. Student characteristics				
Male	0.001	−0.000	0.000	−0.000
	(0.002)	(0.003)	(0.000)	(0.000)
Age at disposition	0.000	−0.001	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.001)
White	−0.031***	−0.004	−0.006***	−0.001
	(0.004)	(0.006)	(0.001)	(0.002)
Black	−0.003	−0.003	0.002**	0.002
	(0.003)	(0.007)	(0.001)	(0.002)
Economically disadvantaged	0.019***	0.006	0.003***	0.002*
	(0.002)	(0.004)	(0.001)	(0.001)
F 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.

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

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^{th} or 10^{th} grade. We standardize these scores to be mean zero and standard deviation one in the full population of test takers in each year. If the student took an exam multiple times in the same academic year, we use their first score. If the student took an end-of-grade and the English I exam in the same year, we use the average of their two standardized scores.

- **Exposure to incarceration (direct and indirect).** We define a student as *directly* exposed to an incarceration if they are linked to a criminal defendant who receives an active sentence; see Appendix C.4 for details on the linking of students and defendants. We define a student as *indirectly* exposed if they attended the same school and grade as a directly-exposed student in the year of the defendant's disposition. Throughout the paper we refer to indirectly-exposed children as "classmates" of the directly-exposed child since these students typically move through the school system together, and thus often share classes.

- **Fighting incident.** A binary indicator for a student having any fighting incident during the academic year, defined by the student offense code 024.

- **Judge stringency.** We define a measure of judicial stringency that captures a judge's average tendency to impose an active sentence among the convictions they oversee (versus an intermediate or community punishment). We use this measure in our judge turnover analysis in Section 3, and also in our analysis of mechanisms in Sections 4–5. Our measure of judge stringency is similar throughout the paper, but the implementation details vary to be appropriate for the two analyses.

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

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

- **Math score.** Math scores include scores on both end-of-grade 3–8 math exams and the end-of-course high school Algebra exam, which is most commonly taken in 9$^{\text{th}}$ grade. We standardize these scores to be mean zero and standard deviation one in the full population of test takers in each year. If the student took an exam multiple times in the same academic year, we use their first score. If the student took an end-of-grade and the Algebra exam in the same year, we use the average of their two standardized scores.

- **Neighborhood SES percentile.** We follow Norris et al. (2021) in defining a measure of mean socioeconomic status (SES) in the neighborhood where the student lives. We define neighborhoods using the student's Census Block Group as reported in the GEOADDRS 20**PUB datasets from NCERDC (see Appendix C.2.2). Since NCERDC collects addresses at the start of the academic year, we define our neighborhood SES measure based on the student's address in the *next* academic year; this allows neighborhood quality to potentially change *in the disposition year* in response to an incarceration.

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

- **Number of suspension days.** The sum of the student's in-school and out-of-school suspension days during the academic year.

- **Offense.** Unless otherwise noted, we define a defendant's offense at the type $\times$ class level. Offense type is either felony, misdemeanor, traffic, or clerk to decide. Felony offense classes are A, B1, B2, C, D, E, F, G, H, I, or clerk to decide. Misdemeanor/traffic offense are A1, 1, 2, 3, or clerk to decide. Offense type $\times$ class is the relevant level for North Carolina's structured sentencing regulations (see Appendix Figures B1–B2). Throughout our analyses, we exclude Class 2 and Class 3 traffic offenses (mainly speeding offenses), as these cases almost never result in an active sentence. In some robustness analyses, we define offenses at the level of a 4-digit offense code (e.g., Misdemeanor Larceny).

- **Student has criminal charge.** An indicator equal to if the student *themself* has a criminal charge in a given year based on our merge of the court (ACIS) and education (NCERDC) datasets (see Appendix C.4). The ACIS data includes each defendant's exact date of birth, while the NCERDC data includes each student's birth month and year. Thus we define the student as having a criminal charge if the defendant that we linked to the student has the same birth month and year as the student. This outcome is defined only for the 2009–2014 academic years because we rarely observe criminal cases with offenses that occurred before before July 1, 2008 or after June 30, 2014. Note that we do not observe criminal activity after the student has left K–12 public school because our merge is based on addresses in the NCERDC data.

C.2 Data

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

C.2.1 Court data

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

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

1. *Case Records (CRCASES).* This dataset includes an observation for each criminal case and includes characteristics of both defendants and their cases. We observe each defendant's race, gender, and age, as well as their exact address at the date of last update in the system. Case characteristics include the date of origination and the court county and type (district/superior).

2. *Offense Records (CROFFNS).* This dataset includes an observation for each charge of each criminal case, with information on the offense and disposition outcome. These records include the date of arrest, arraignment, and disposition, as well as a separate 4-digit offense

code for each of these three events.[34] 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 contains information on the schools students attended during the year, their grades, and days of absence. They also include demographic characteristics such as age, race, and gender. We use these datasets to define the full population of students in public school in a given year.

2. *Current Test File (CURTEST_PUB20**).* These datasets include information on student exam performance with a separate observation for each student/year/test. We use the grade 3–8 end-of-grade reading and math scores, and the end-of-course Algebra and English I exams

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

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.[35]
    - Latitude and longitude of 2010 Census blocks.[36]
    - Data on child incarceration outcomes by Census tract from the Opportunity Insights project.[37]

[35] Obtained in November 2020 from: https://data.census.gov/cedsci/.
[36] Obtained in February 2021 from: https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2010.html.
[37] Obtained in February 2021 from: https://opportunityinsights.org/data/.

- A crosswalk between 2000 and 2010 Census blocks.[38]

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

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

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

C.2.3 Prison data

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

Our main dataset from the prison records is the *Sentence Components (OFNT3CE1)*. This dataset contains information on each component of each prison sentence as committed by the court system. These records cover court cases ranging as far back as the 1970s up through 2016. The variables include the case number (`cmcaseno`) and county (`cmcocnvt`), which we use to reproduce the case identifier in the court data (`crrkey`) to merge the court and prison records. Our main outcome from this dataset is an indicator for having any active prison sentence resulting from a given case, which is defined by commitment prefixes (`cmprefix`) that begin with letters.[39]

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

[39] Probation sentences have commitment prefixes that begin with numbers.

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

C.3 Court data cleaning

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

C.3.1 Creating a defendant panel

In the ACIS data, the most granular unit of observation is a charge (`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.[40] This identifier is based on the defendant identification variables that we observe in the court data: full name (`crrnam`), date of birth (`crrdob`), driver's license number (`crrdln`), last four SSN digits (`crrssn`), and street address (defined by `crradd`, `crrcty`, `crrdst`, and `crrzip`). We assign the same ID number to observations that match exactly on full name and at least one of the other four identification variables. To allow for typos or variations in the recorded name, we also use a fuzzy match of names in combination with the requirement that observations match on at least two of the other identification variables.

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

[40] Creating a defendant ID also allows us to examine defendant recidivism outcomes.

a defendant/disposition pair, we keep the observation with the most serious 4-digit offense code at the time of arrest (`croffc`). We define the most serious offense as the one that has the highest offense class.[41] 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 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.[42] Cases are almost always

[41] Specifically, the offense class ordering is Felony A $>\cdots>$ Felony I $>$ Misdemeanor A1 $>\cdots>$ Misdemeanor 3 $>$ Traffic A1 $>\cdots>$ Traffic 3 $>$ Infraction.

[42] The case's county is indicated by the first three digits of the case number `crrkey`. Case types are defined by the

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

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

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

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

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

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

---

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

C.4 Merge

To link our court and education records, we use a crosswalk between exact addresses in the two datasets that was created by the NCERDC. We sent the NCERDC a dataset with the address variables that are available in the court records, which include street address (`crradd`), city (`crrcty`), state (`crrdst`), and zip code (`crrzip`). The sample for this dataset included both criminal cases and infractions from July 1, 2009 through June 30, 2014, as described in Section C.2.1 above. The NCERDC linked these variables to the geocoded address identifier available in their records (`geo_addrid`) using their confidential information on exact address. We received back from the NCERDC a dataset that provides a crosswalk between `geo_addrid` in the education data and the address variables in the court data.[43]

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

Panels B–D of Appendix Table C1 provide evidence that NCERDC's merge of addresses was high quality. The proportion of addresses that are in the largest counties (Panel B) and cities (Panel C) is broadly comparable across the datasets and samples, although these proportions also vary for reasons unrelated to merge quality (e.g., criminal activity and sorting of households with children). More convincingly, Panel D shows that the vast majority of defendant and student addresses in our linked sample have the same county and Census Block Group. To compute this statistic, we geocoded defendant addresses in the ACIS data to obtain the Census Block Group of each address, as this variable also appears in the NCERDC data.[44] Among addresses where these

[43] 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.

[44] The notes to Appendix Table C1 provide details on our geocoding of defendant addresses.

TABLE C1: Merge statistics for NCERDC and ACIS addresses

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

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

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

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

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

Using the address crosswalk, we link students and defendants who live at the same address in the same year. Specifically, to attach a student to a defendant, we require that the student's reported address in a given academic year matches the year of last update of the defendant's case in the court data. Addresses in the NCERDC are collected at the beginning of the academic year; for example, addresses for the 2013–2014 academic year were based on where students lived in mid-2013. These students are linked to defendants who have the same address and for which the date of last update is between January 1, 2013 and December 31, 2013.

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

A second and more important source of potential mismatch arises from apartment buildings and other multi-unit addresses. Most addresses in the court data do include information on the unit number if it exists, but the geocoded addresses in the NCERDC data (`geo_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.

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

C.5 Samples

Our paper uses five different samples to examine the impacts of community-level exposure to incarceration and its mechanisms. In Section 3, we use all North Carolina public school students for our judge turnover strategy (column (A) of Table 1). In Section 4, we use two samples that are appropriate for our event study and judge stringency strategies to examine the direct effects of exposure to household incarceration (columns (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

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

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

Notes: This table compares the race/ethnicity of defendants and students who we linked based on the merge process described in Appendix C.4. Panel A includes all linked defendants and students, which is the sample reported in column (C) of Table 1. Panel B includes students in our direct exposure event study sample, which is the sample reported in column (A) of Appendix Table A3. If the student was linked to multiple defendants, we use the defendant with the most serious offense as defined in Appendix C.5.

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

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

C.5.2 Direct exposure samples

For Section 4, we define two samples of students who were directly exposed to a household incarceration: one for our event study strategy, and another for our judge stringency strategy. These samples differ because of the different sources of identification. Nonetheless, the students in these two samples have very similar characteristics (columns (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.[45] We exclude defendants that never received a guilty verdict (e.g., if their case was dismissed or if they were found not guilty). We also exclude defendants whose most serious event involved a Class 2 or Class 3 traffic offense (mainly speeding offenses), as these charges almost never result in an active sentence.

Next, we restrict to students and defendants who are likely to be related, and we identify the defendant with the most serious court event among those that we linked to a given student. Our merge of students and defendants is at the address level (see Appendix C.4), which means that students who live in multi-unit buildings may not necessarily be related to the defendant. To improve the quality of matches, we exclude any student that is linked to more than 3 defendants, and we exclude any defendant that is linked to more than 6 students.[46] If the student is linked to multiple defendants, our event study sample uses only the defendant with the most serious court event (defined by the same criteria as in the previous paragraph).

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

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

[45] 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).

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

umn (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.[47] We exclude cases that never faced a judge or that never received *any* verdict; these are primarily cases that were dismissed by the district attorney or hearings for probation violations. As above, we also exclude cases where the most serious charge was a Class 2 or Class 3 traffic offense.

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

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

The resulting judge stringency sample includes 118,416 students linked to 79,765 defendants, as shown in column (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

[47] 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.

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 A4). Our regressions in Section 5 are at the directly-impacted student $\times$ classmate $\times$ year level, so classmates often appear in the regressions multiple times because they are linked to multiple students in the direct exposure samples.