Stopped by the Police: The End of "Stop-and-Frisk" on Neighborhood Crime and High School Dropout Rates*

Jonathan Tebes†
University of Notre Dame

Jeffrey Fagan‡
Columbia University

December 2022

Preliminary draft for 2023 ASSA Meeting

Abstract

Over 3.5 million pedestrians are stopped by police in the United States every year. This paper explores the effectiveness of using pedestrian stops as a crime deterrence tool. Using administrative data from New York City, we test whether the concentration of pedestrian stops in higher-crime neighborhoods deters neighborhood crime and whether frequent exposure to police stops affects dropout rates of neighborhood high school students. Exploiting a 2012 reform that reduced stops by 95%, we compare neighborhoods that have similar crime rates but substantially different stop rates prior to the reform. Treated neighborhoods that experienced twice the reduction in stop rates do not display differential increases in felonies and violent misdemeanors, shootings, or killings over the five years following the reform. Analysis of police surges, however, indicates that when increased stop rates are accompanied by an increase in patrol officers, serious crime declines. But alone, heightened stop rates have no measurable impact on serious crime. Comparing students across schools that are differentially exposed to changes in stop rates, we estimate that the reform reduced the probability of high school dropout by about 660 students per academic year, carrying an annual social value of over $205 million.

Acknowledgements: We thank Edward Glaeser, Lawrence Katz, Joscha Legewie, and Andrei Shleifer for unrelenting advice and encouragement on this project. We are grateful for helpful comments from Desmond Ang, Noam Angrist, Caroline Chin, Natalia Emanuel, Paul Fontanier, Benny Goldman, Dev Patel, Robert Fluegge, and seminar participants at Harvard Public/Labor lunches and seminars. The NYU Research Alliance of NYC Schools provided the education data for this project. Tebes also benefited from generous financial support from the NSF Graduate Research Fellowship, the Stone Scholar Fellowship for Social Policy and Inequality Research, and the Horowitz Foundation. Fagan was instrumental in securing and understanding NYPD policing data. All errors are our own.

†Department of Economics, University of Notre Dame, Notre Dame, IN 46556. jtebes@nd.edu.

‡Columbia Law School, Columbia University, New York, NY 10027.

I Introduction

In the aftermath of George Floyd's murder, 20 million Americans protested police brutality and systemic racism (Buchanan et al., 2020). Many of them called for an overhaul of existing police institutions. Yet city-wide increases in officers per capita have consistently been found to causally reduce serious crime, leading some scholars to even claim U.S. cities are currently under-policed (Evans and Owens, 2007; Chalfin and McCrary, 2018; Mello, 2019). Much less is known about the effectiveness of pedestrian stops, although over 3.5 million pedestrians are stopped by police in the United States every year (Harrell and Davis, 2020). While the primary goal of pedestrian stops is to protect civilians from imminent harm, police departments often concentrate stops in higher-crime neighborhoods with the aim of deterring serious crime.

The impact of pedestrian stops on neighborhood crime and community well-being is theoretically ambiguous. Increased stop rates could raise the probability of apprehension and thus reduce the expected payoff of criminal activity (Becker, 1968). Alternatively, stops may impose substantial costs on local communities, as stops can be traumatizing (Geller et al., 2014; Boyd, 2018), breed institutional distrust (Kirk and Papachristos, 2011), and disrupt educational investments (Legewie and Fagan, 2019; Bacher-Hicks and de la Campa, 2020). Stops also mechanically increase the likelihood of more serious downstream police actions, such as uses of force or arrests for minor offenses (Knox and Mummolo, 2020).

This paper explores the effectiveness of using pedestrian stops as a crime deterrence tool. To this end, we analyze the impact of a 2012 federal lawsuit – *Floyd, et al. v. The City of New York, et al.* – that ruled NYPD's stopping practices were unconstitutional and led to a permanent, 95% reduction in city-wide stop rates. We assess *effectiveness* by estimating the impact of the reform on two key outcomes: neighborhood crime and high school dropout rates. We complement our crime analysis with an analysis of police surges that increased both officers and stops in higher-crime areas prior to the reform, allowing us to disentangle the effect of stopping protocols from patrol officer presence.

We begin the crime analysis by first showing that city-wide crime trends appear unaffected by the reform. Aggregate crime responses, however, may mask a reduction in latent criminal activity occurring simultaneously or heterogeneous neighborhood responses. To estimate the causal effect of stops on neighborhood crime, we identify neighborhoods that, prior to the reform, have similar crime rates but differ substantially in stop rates. Specifically, we split neighborhoods into treatment and control groups based on mean stop rate residuals from a regression of stop rates on neighborhood crime measures during an earlier training period. Using a flexible difference-in-differences framework, we explore how crime rates differentially evolve in treatment neighborhoods relative to control neighborhoods over the five years following the reform. For the four years prior to the reform, treatment and control neighborhoods display parallel trends for multiple crime and policing measures.

Difference-in-differences estimates reveal that stop rates, uses of force, and frisks all fall by twice as much in treatment relative to control neighborhoods. Treatment neighborhoods, however, do not display differential increases in felonies or violent misdemeanors, major felonies, shootings, or killings over the five years following the reform. The confidence interval on felonies and violent misdemeanors rules out an increase of 1.5% of the pre-period mean. To put this null effect in context, New York City experienced a 38% decline in felony crime from 2000 to 2010. Counter to the crime deterrence hypothesis, we find that treatment neighborhoods exhibit *reductions* in non-violent misdemeanors and violations, which can be entirely explained by reductions in stop-related arrests.[1]

Given the lack of crime responses to the reform, we investigate whether police patrols, more broadly, deter crime. To do so, we analyze the impact of neighborhood police surges prior to the reform. About every six months, the police commissioner selected "Impact Zones" to receive additional patrol officers and resources. Using a difference-in-differences framework that controls for linear pre-trend differences, we find that Impact Zone assignment increases both stop rates and the number of officers conducting stops by about 35%. By year three of assignment, we find that felonies and violent misdemeanors are reduced by 8% ($p = 0.011$) and major felonies are reduced by 11% ($p = 0.015$). These results are robust to allowing for reasonably large parallel trend violations a la Roth and Rambachan (2021), and to a broader geographical definition of Impact Zone exposure that takes into account potential crime spillovers on nearby neighborhoods. Given that officer assignments were relatively unaffected by the reform, our combined findings imply patrol officer presence but not stop rates matter for deterring serious crime.

Employing a similar differences-in-differences strategy, we next use the reform to estimate spillovers of neighborhood stop exposure on high school dropout rates.[2] We proxy for stop exposure in a student's peer network by exploiting school-level variation in stop exposure. The advantage of this approach is that it is able to detect effects in the presence of within-school peer effects and accounts for the fact that high school students likely spend time in areas outside of their home Census tract. Specifically, we split schools into quartiles based on the number of stops per square mile observed in students' home Census tracts during the three school years prior to our sample period. We then estimate differential changes in yearly

[1] Our results are robust to alternative treatment definitions, controls, and parallel trends violations a la Roth and Rambachan (2021).

[2] We focus on high school students since they are 21 times more likely to be stopped by police than middle school students and face critical educational investment decisions.

enrollment outcomes for students who attend fourth and third quartile schools relative to students attending "control" schools that rank in the bottom half of training-period stop exposure.

As a result of the reform, students attending fourth (third) quartile schools experience an annual decline in stops in their home neighborhoods that is more than three (two) times larger than that same decline for students in control schools. We measure the effect of the reform using two educational outcomes. The first outcome is an indicator for whether a student was discharged by a non-DOE agency directive ("institutional directive"), such as from the Department of Corrections. This provides a lower-bound estimate of criminal justice-related school dismissals. The second outcome is an indicator of whether or not a student drops out of high school or is discharged by an institutional directive. This outcome measures the holistic impact of the reform on the likelihood a student leaves high school in a given school year.

We find that students in fourth-quartile schools display a 0.14 percentage point or 54% reduction in likelihood of being discharged by an institutional directive relative to students attending control schools ($p = 0.004$). That is, 84 students per year were *not* discharged by institutional directive following the *Floyd* decision. Depending on the specification, we estimate that the likelihood that a fourth-quartile student dropped out or was discharged by institutional directive fell by 0.36-1.66 percentage points per school year ($p < 0.02$), preventing about 660 students per year from leaving high school prematurely. IV estimates indicate that increasing the average stop rate in students' home neighborhoods by 100 stops per year carries an annual social cost of $23 million. Moreover, our results likely underestimate true effect sizes since impacts on students in control schools are differenced out.[3] Both in terms of student-level effect sizes and the number of students affected, effect sizes are substantially larger than previous estimates for middle school students as well as estimated impacts of neighborhood exposure to police killings (Ang, 2021; Bacher-Hicks and de la Campa, 2020). Heterogeneity analyses reveal that Black and Hispanic male students are the most likely to benefit from the reform.

This paper makes three main contributions. First, we contribute to an understanding of crime deterrence.[4] Our analysis of the reform provides new causal evidence on the effectiveness of pedestrian stops as a crime deterrence tool. The precision of our null result suggests that, relative to alternative policing activities, pedestrian stops have very little impact on serious crime. These findings reinforce the notion that proactive and disorder

[3]We confirm that our results are not driven by other educational reforms over this period, such as changes to suspension policies.

[4]See Chalfin and McCrary (2017) for a thoughtful summary of criminology and economic literature on this subject.

policing tactics, which aim to deter crime by strictly enforcing low-level offenses and maintaining order in higher-crime areas, have only limited impacts on serious crime (Braga and Bond., 2008; Caetano and Maheshri, 2014; MacDonald et al., 2016; Chalfin and McCrary, 2018). On the other hand, analysis of Impact Zones accords with a large literature that finds hot-spot policing (Braga and Bond., 2008; Weisburd et al., 2006; MacDonald et al., 2016), officer deployments (Di Tella and Schargrodsky, 2004), and police employment (Chalfin and McCrary, 2018; Mello, 2019) to effectively deter neighborhood crime.

Second, we add to a growing body of evidence that finds policing to impose negative externalities on local communities of color (Mello, 2021; Ang, 2021; Legewie and Fagan, 2019; Bacher-Hicks and de la Campa, 2020). Over 90% of the 660 students who were prevented from dropping out of high school by the *Floyd* decision were Black or Hispanic. Previous work finds that exposure to aggressive policing during middle school can harm contemporaneous test scores, high school graduation rates, and college-going rates of Black male students (Legewie and Fagan, 2019; Bacher-Hicks and de la Campa, 2020). This paper studies stop exposure during high school – a time when stops are 21 times more likely than during middle school. By leveraging sharp changes in city-wide stop rates, we are better able to gauge the full magnitude of treatment effects, and have the advantage of studying a reform that has the potential to be more broadly adopted.

Lastly, this paper sheds light on the mechanisms underlying the impact of neighborhoods on social mobility (Ludwig et al., 2013; Chetty and Katz, 2016; Chetty et al., 2020). Chetty et al. (2020) show that, conditional on parental income, Black boys have lower incomes in adulthood in 99% of Census tracts. In companion work, we show that Black male teenagers are substantially over-stopped by police, while this paper shows that these interactions disproportionately translate into high school dropouts and interactions with the carceral state. Combined with previous studies showing the indirect harm of police killings (Ang, 2021), ticketing (Mello, 2019; Goncalves and Mello, 2021), and prosecution of low-level arrests (Agan et al., 2021a), it becomes clear that the police play a pivotal role in the life trajectory of young Black men. This paper thus provides actionable policy advice about one reform that could help close place-based racial opportunity gaps.

The remainder of this paper proceeds as follows. Section II describes the institutional details and provides descriptive statistics. Section III describes data sources and outcomes. Section IV estimates crime deterrence effects. Section V estimates impacts on high school dropout rates and Section VII concludes.

II Setting

A Pedestrian stops

Officers are afforded considerable discretion over whether or not to conduct a stop. They may stop a pedestrian if there is a *reasonable suspicion* that he has committed or is about to commit a crime (*Terry v. Ohio*, 1968). That is, the officer must be able to articulate specific facts that, together, would lead a reasonable person to suspect criminal activity. These include fitting the description of a suspect near the location where a crime was reported, walking in a way that evades police contact, carrying items that have been reported stolen recently, or wearing bulky clothes in warm weather. If the officer believes the pedestrian to be armed and dangerous, they may "frisk" or pat down the pedestrian's outer clothing to search for a weapon. During the course of a stop, if the officer observes evidence of criminal behavior, such as unlawful possession of drugs or a weapon, "probable cause" is established and the officer may arrest the pedestrian.

Investigative stops primarily serve two policing functions. First, they empower officers to protect civilians from imminent harm by allowing them to investigate suspicious activity. Stops can also be particularly useful for confiscating illegal firearms before they are used in violent crimes, although our data show firearms are rarely recovered. Second, the concentration of stops in higher-crime areas may deter future crime by increasing the probability an offender is apprehended (Becker, 1968). In fact, it is exactly this deterrence rationale that Mayor Bloomberg employed to justify racial disparities in stop rates during the stop-and-frisk era:

> *They have argued that police stops are discriminatory because they do not reflect the city's overall census numbers. By that flawed logic, our police officers would stop women as often as men and senior citizens as often as young people... The absurd result of such a strategy would be far more crimes committed against black and Latino New Yorkers. When it comes to policing, political correctness is deadly.* (Bloomberg, 2013)

The use of pedestrian stops increased dramatically in the 1990s, as urban police departments transitioned from "reactive" to "proactive" policing strategies. The NYPD was at the forefront of this transition, adopting Compstat in 1995, which was the first crime tracking and management system in the U.S. CompStat enabled the police to allocate officers to neighborhoods known for crime during peak hours of criminal activity. Commanders attended monthly CompStat meetings designed to hold them accountable to reducing major felony crimes in their designated area. As an advocate of "broken windows" policing theory – the idea that visible signs of crime and civil disorder encourage more serious crimes – Commissioner Bratton instructed commanders and their line officers to conduct a high volume of stops and strictly enforce minor offenses, such as fare evasion, public drinking, and graffiti. During the 1990s, homicide rates fell in New York City by 74%, the largest decline of any large city over this period (Levitt, 2004). Interpreting this decline as evidence of efficacy, other departments quickly created their own data-driven management systems and adopted similar proactive policing strategies.[5] Data-driven, proactive policing would become a prominent fixture of urban policing strategies for the next two decades (National Academies of Sciences Engineering and Medicine, 2018).

From 2006 to 2011, the NYPD conducted over 545,000 stops per year, a rate that is 10,000 times greater than the annual number of police-involved shooting incidents.[6] As shown in Figure I, stops were concentrated in higher-crime neighborhoods that were disproportionately Black and Hispanic – 84% of stops were of Black or Hispanic pedestrians compared to a residential population that is just 51% Black or Hispanic. Much of this racial disparity can be explained by where stops were conducted. Assigning the residential racial composition of a Census tract to stops conducted in that tract predicts that 69% of stops would be Black or Hispanic, suggesting that over half of observed racial differences in stop rates can be attributed to the concentration of stops in predominantly Black and Hispanic neighborhoods.

Pedestrians stopped by police were also disproportionately male (93%) and young – 54% were under the age of 25, and a third were under the age of 20.[7] Using residential population estimates from the 2010 Census, we estimate that in higher-crime neighborhoods, the police conducted 7.5 stops per Black male resident between the ages of 14 and 18 per year.[8] For Hispanic males, the equivalent stop rate was 2.8 stops per resident, while it was only 0.4 stops for male teenagers of other races. This means that during high school, the average Black (Hispanic) male student living in a higher-crime neighborhood was stopped by police 30.0 (11.2) times. Most of these stops did not result in the detection of criminal activity. Of the 14% that did detect pedestrian wrong-doing, 5.8% resulted in an arrest, 0.9% led to the discovery of a weapon, 1.7% led to the discovery of drugs, and 6.4% led to a court summons being issued for a minor violation. Only 1 in 770 stops led to the discovery of an illegal

[5]Five years after its creation, one-third of the nation's 515 largest police departments intended to deploy a CompStat-like program by 2001 (Eterno and Silverman, 2019).

[6]We divide the number of stop incidents by the number of officer-involved shooting incidents recorded by the NYPD from 2006-2011.

[7]To see the full distribution of stops by pedestrian age, see Figure A.2.

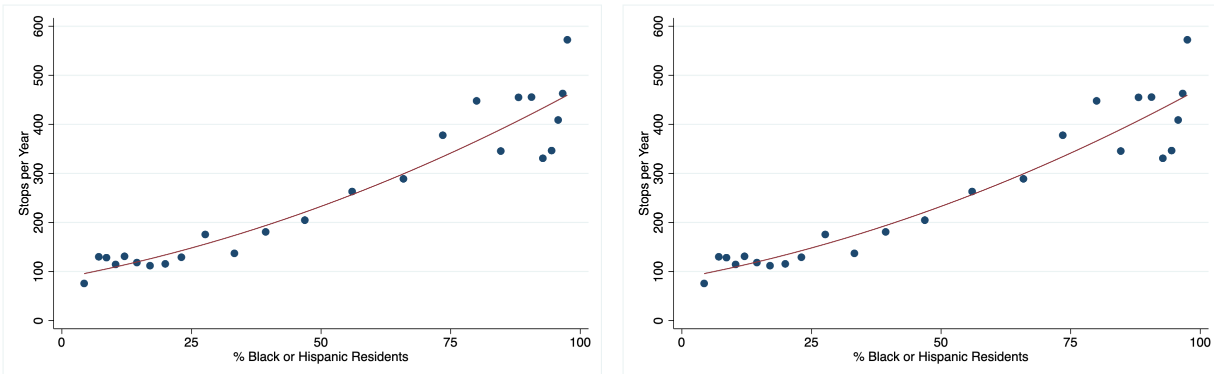
[8]Higher-crime neighborhoods are defined as tracts that rank in the top 25% of felony crimes per square mile during the training period. We estimate race-specific 14-18 year-old male populations for each Census tract by computing a weighted average of 10-14 and 15-19 year old bins and multiplying these estimates by the estimated racial composition of the entire Census tract.

firearm.

Figure I: Neighborhood Stop Rates by Felony Crime Rates & Racial Composition in NYC

(a) Neighborhood Felony Crime Rate

(b) Neighborhood Racial Composition



Notes: This figure displays the number of stops per year by neighborhood characteristics. The x-axis of Panel A refers to the percentile rank of each tract in terms of felony crimes per per 1 mile^2 recorded between Jan 1, 2006 and Apr 15, 2008. The x-axis of Panel B refers to the fraction of residents who identify as Black or Hispanic in the 2010 Census. The sample includes 2,096 of 2165 Census tracts, dropping tracts with less than 500 residents listed in the 2010 Census. Sample period spans Jan 1, 2006 to Apr 15, 2012.

Surprisingly, stop rates recorded during the height of stop and frisk are similar to stop rates in *other large urban police departments over the past five years.* Table I compares stops and stop outcomes across New York City, Chicago, and Philadelphia – the only departments of the ten largest police departments to publicly release information on all pedestrian stops.[9] Annual stop rates in Philadelphia and Chicago were, respectively, 135% and 90% the level conducted by the NYPD from 2006-2011. After the reform, stop rates in New York City fell to just a fraction of what stop rates are currently in Philadelphia and Chicago. The reform we study is therefore relevant for many other large U.S. cities.

B ***Floyd et al. v. The City of New York***

This paper primarily explores the consequences of a federal lawsuit – *Floyd, et al. vs. City of New York, et al.* (hereafter referred to as "*Floyd*") – that alleged the NYPD implemented and sanctioned "a policy, practice, and/or custom of unconstitutional stops and frisks of City residents" in violation of the 4th and 14th Amendments of the U.S. Constitution. This was one of the first rulings to severely limit the use of investigative pedestrian stops. Since

[9]Most police departments do not release information on pedestrian stops. Of the ten largest police departments in the country, only four release any data on investigative stops. Only three provide information on all stops, including stops that do not result in an arrest or citation. All three departments released these data publicly in order to comply with a court mandate.

Table I: Pedestrian Stop Rates in New York City, Chicago, & Philadelphia

	NYPD		Chicago PD	Philadelphia PD
	Pre-Floyd	Post-Floyd		
	(1)	(2)	(3)	(4)
Stops	546,745	19,001	161,755	141,506
Stops per 100k	6,625	223	5,959	9,008
Stops per Violent Indexed Crime	11.07	0.39	5.74	9.38
Stops per All Indexed Crime	2.81	0.11	1.41	2.22
Stops per Full-time Officers	15.55	0.54	12.78	21.96
Stops: % Black or Hispanic	83.56	85.35	90.22	78.79
$\frac{\text{\% of Stops Black or Hispanic}}{\text{\% of Population Black or Hispanic}}$	1.63	1.68	1.56	1.42
Stops: % with Arrest Made	5.79	24.05	13.13	11.43
Stops: % with Weapon Discovery	0.95	6.69	2.04	-
Stops: % with Gun Discovery	0.13	1.91	1.38	-
Stops: % with Drug Discovery	1.73	6.70	6.13	3.40
Stops: % with Court Summons Issued	6.44	2.80	7.97	-
Years	2006 - 2011	2014 - 2019	2016 - 2019	2014 - 2019

Notes: This table provides descriptive statistics on stops conducted by the New York City Police Department (NYPD), the Chicago Police Department (CPD), and the Philadelphia Police Department (PPD). Data for the CPD and PPD come from administrative records that are publicly available on each department's website. Crime data come from the FBI's Uniform Crime Reporting (UCR) program. All index crime refers to the sum of violent and property crime as reported in UCR data. The fraction of a city's population that is Black or Hispanic is measured using the 2010 Census for Column 1 and the 2019 ACS 5-year estimates for Columns 2 through 4.

this case, the ACLU and the Department of Justice have taken legal action to implement similar reforms in Chicago, Ferguson, and Philadelphia (ACLU of Illinois, 2015; Dept. of Justice, 2016; ACLU of Pennsylvania, 2011, 2021).

We display a timeline of important moments in the case and changes to the NYPD's stop and frisk policies in Figure A.1. Even though the case was first filed in January 2008, the NYPD continued to increase annual stop rates through the beginning of 2012. In fact, an internal memo issued by Commissioner Kelly in October 2011 recommended managers set performance goals for the number of monthly pedestrian stops conducted by patrol officers. On April 16, 2012, the Court allowed expert testimony from Dr. Fagan, which indicated that the vast majority of stops did not result in detection of any wrong-doing and that Black and Hispanic residents were stopped at disproportionate rates. A month later, the court granted "Class Action" status, allowing potentially thousands of stopped pedestrians to join the lawsuit. That same day, Commissioner Kelly issued a memo that formalized efforts to scale back stop rates, which had already begun to decline following Dr. Fagan's testimony. Around this time, Mayor Bloomberg released a brief statement supporting the reform, stating: "we need to mend, not end, the practice, and the reforms Commissioner Kelly has put into place ensure the focus is quality, not quantity." (Goldstein and Ruderman, 2012) [10] After a nine-week trial, on August 13, 2013, the court found the NYPD liable for a pattern and practice of racial profiling and unconstitutional stops, and ordered broad reforms that would be supervised through a court monitor.

The monitor implemented three key reforms. First, the police academy added training to help cadets distinguish between lawful "criminal profiling" and illegal "racial profiling". Second, the police revised its patrol guide to more clearly and narrowly define when officers have lawful authority to conduct a stop under *reasonable suspicion*. Third, the monitor worked with the NYPD to improve data collection and management processes to ensure officers were keeping an honest record of all stops (Zimroth, 2016).

Figure II plots city-wide trends in stops, major felonies, non-major felonies and violent misdemeanors, and non-violent misdemeanors and violations throughout the study period. The first dashed line is set to the week of April 16, 2012, since this is when the NYPD began reducing stop rates in response to the case. The second dashed line marks the court's decision. Panel A illustrates that weekly stops fell by 95% over this period. Panel B, C, and D illustrate that reported crime rates remained rather flat for the first few years following the reform before declining slightly. Importantly, there is no evidence that aggregate crime rates rose in response to *Floyd*.

There is little evidence that changes to other aspects of policing coincided with the reform. The assignment of additional officers to high-crime areas under Operation Impact (e.g. see below) continued through July 2015, more than three years after the NYPD started to reduce stop rates. Similarly, the number of officers employed by the NYPD and city-wide clearance rates remained relatively flat for over two years after the reform (see Appendix Figure **??** and **??**). Finally, we do not observe a break in trend for total pay or overtime pay following the reform (see Appendix Figure **??**). Given this context, we interpret the reform as a shock to stopping protocol that held officer assignments relatively fixed.

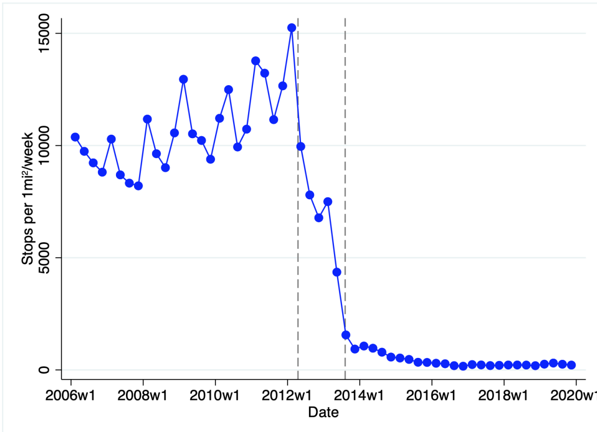
With the reduction of stops, patrol officers were able to shift their attention to other patrol activities, such as pursuing reports of more serious crime or building community ties. Broad adoption of community policing tactics, however, would not be adopted for over three years after Dr. Fagan's testimony.[11]

[10] In January 2013, the same court ruled in a separate lawsuit – *Ligon et al. v. The City of New York* – that the NYPD must immediately cease its practice of unlawful trespass stops outside public housing buildings in the Bronx, further reducing stop rates.

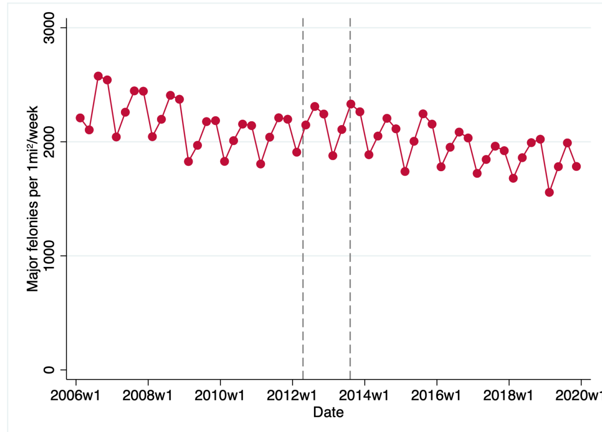
[11] In June of 2015, Commissioner Bratton established Neighborhood Coordinating Officers who were tasked with spending time each day engaging community members and nurturing relationships in their sector.

Figure II: City-wide Trends during Study Period

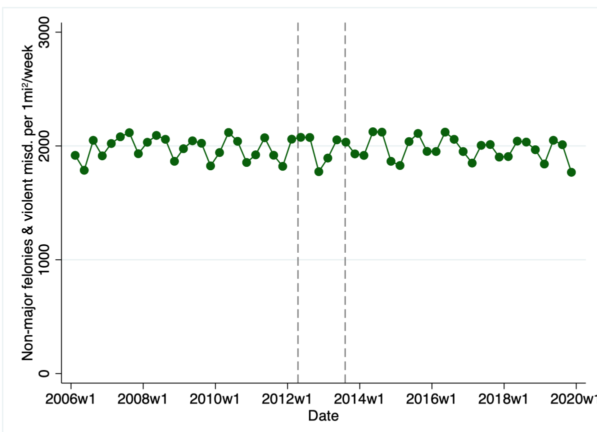
(a) Stops



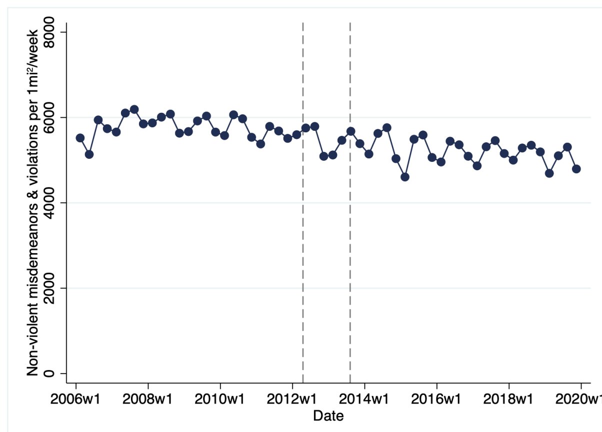
(b) Major Felonies



(c) Non-Major Felonies & Violent Misdemeanors



(d) Non-violent Misdemeanors & Violations



Notes: This figure graphs city-wide means in stops, reported major felonies, non-major felonies and violent misdemeanors, and non-violent misdemeanors and violations by quarter from 2006 through 2019. All outcomes are measured as weekly rates per 1 mile^2. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling.

C Operation Impact

Operation Impact began in 2003 as a "natural outgrowth" of CompStat (Golden and Almo, 2004). About every six months, precinct and borough commanders nominated crime hot-spots that the police commissioner would then select to be "Impact Zones" to receive additional officers and resources. In the first phase of Operation Impact, about 1,500 officers or two-thirds of all new academy graduates were deployed to Impact Zones. There were four general policing strategies employed in Impact Zones. The first was to increase foot patrols during times when crimes are most commonly committed. This included “vertical patrols”, where officers would sweep the halls of public housing units, top-to-bottom. Second, officers were instructed to issue citations and court summonses for low-level “quality-of-life” crimes, such as loitering or drinking alcohol in public. Third, officers were encouraged to stop and question anyone they suspected of committing crimes in the area. Finally, the department allocated more specialized units, such as the Firearms Investigation Unit, to address the particular crime problems of the Impact Zone. (Golden and Almo, 2004)

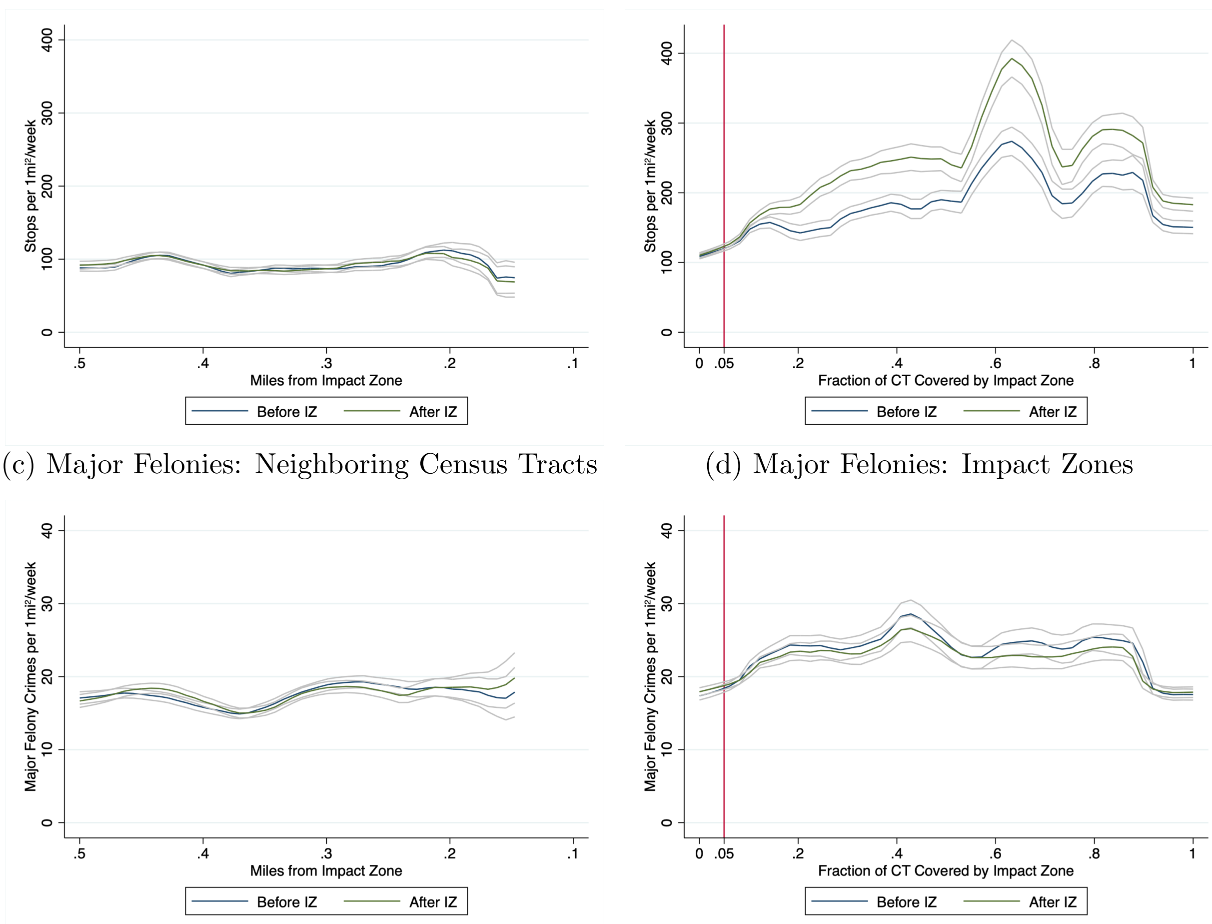
Figure III: Impact Zone Assignment on Stops and Major Felonies

(a) Stops: Neighboring Census Tracts

(b) Stops: Impact Zones

(c) Major Felonies: Neighboring Census Tracts

(d) Major Felonies: Impact Zones



Notes: This figure displays local polynomial regressions of stops and major felony crimes per square mile by Impact Zone proximity, separately for before and after an Impact Zone assignment event. All outcomes are residualized by week-year fixed effects and then added back to the constant term of the fixed effect regression. The estimation sample consists of tract-week observations for the 23 weeks surrounding an assignment event, since this is the minimum duration of all assignment periods. Panels B and D include neighborhoods that, following assignment, are partially covered by an Impact Zone, while Panels A and C include neighborhoods within 0.5 miles of a Impact Zone. Standard errors are estimated using pilot bandwidths equal to 1.5 times the kernel bandwidths and are depicted as grey bandwidth lines (Fan et al., 1996).

We obtained maps of Impact Zone borders from the NYPD for thirteen phases that span January 2006 through July 2012. Overlaying Census tracts with these maps, we generate a panel dataset containing information on the fraction of a Census tract's area that is covered by an Impact Zone as well as the distance from an Impact Zone for tracts that are not touched by an Impact Zone.[12] In a fashion similar to Ang (2020), we plot the distance gradient of changes in stops and major felonies for the five months before and after an assignment event.[13] Figure III shows changes in stops and major felonies for tracts that prior to an assignment event were greater than 0.5 miles from an Impact Zone but, following assignment, were partially covered by an Impact Zone (Panels B and D) or were within 0.5 miles (Panels A and C). Panels A and C demonstrate that both stops and major felonies were initially unaffected by the assignment of a neighboring tract to an Impact Zone. Panel B demonstrates that stop rates begin to diverge when 10% or more of a tract is covered by an Impact Zone, and reach about 35% of the pre-assignment mean at 20% coverage. To be conservative, we define Impact Zone treatment as a tract with over 5% of its area covered by an Impact Zone.

III Data

A NYPD Data on Policing and Crime Measures

Policing and crime data primarily come from administrative NYPD records. These data include information on pedestrian stops from the "Stop, Question, and Frisk" program, information on criminal activity compiled from crime *complaints*, arrest records, and shootings data, and boundary maps on the roll-out of Operation Impact.

Stops were recorded by NYPD officers on the "Stop, Question, and Frisk Report Worksheets", known as form UF-250, and contain geographic coordinates of where a stop occurred, date and time-of-day information, demographic information on the person stopped (i.e., race, sex, approximate age), information on police actions taken (i.e., frisk, searched, used force and type of force used), and stop outcomes. Stop outcomes include whether the stop led to an arrest, the discovery of drugs or weapons, or the issuance of a court summons. We use proprietary stop data obtained by Dr. Fagan for years 2006-2016, since these data addition-

[12]To ease computation, we measure distance as the minimum distance from the centroid of a Census block within an uncovered Census tract to a Census block within a "covered" Census tract.

[13]Specifically, we include the 23 weeks before and after Impact Zone assignment since all assignment windows in our study period have a duration of at least 23 weeks.

ally contain de-identified officer IDs describing the lead officer who conducted the stop.[14] For years 2017-2019, we use publicly-available stop data.

Prior to 2012, administrative stop records were used internally to monitor whether officers were conducting a sufficiently high number of pedestrian stops, and thus under-reporting is of limited concern. While some of the observed reduction in stop rates due to *Floyd* may have resulted from under-reporting, it is likely this fraction is small. Following *Floyd*, the court-ordered monitor worked closely with NYPD brass to monitor and audit officer reporting of stops.[15] Second, if the majority of changes in stop rates were due to under-reporting, we would expect to see hit rates rapidly increasing after *Floyd*, since then the only recorded stops would be those that detected criminal activity. However, this is not the case; hit rates remain surprisingly flat throughout the first year post-*Floyd*, as stop rates fell by over 35%.

Crime data capture all crime complaints reported to the NYPD since 2006, including incidents that were never closed by police. Data are at the incident-level and include information on the date the complaint was reported to police, geographic coordinates of where the incident took place, and offense categories that describe the most serious offense related to an incident. Offense information includes broad severity categories, such as whether the crime was a "felony", "misdemeanor", or "violation", as well as more granular offense types (e.g., murder, rape, felony assault, grand larceny, etc.). Crime reports can be initiated by civilians *or police officers* that observe criminal activity. To this end, police presence in a neighborhood can affect observed crime rates by altering civilian reporting and police detection of criminal activity.

We supplement crime reports with administrative records on all arrests and reported shooting incidents that occurred in NYC since 2006. Both datasets contain information on the date, time-of-day, and geographic coordinates of each incident. Arrest data contain the same detailed offense categories recorded in crime complaints.[16] Unfortunately, we cannot link crime reports to arrest records, and thus cannot observe how clearance rates vary across neighborhoods, as arrests are frequently made outside of the incident neighborhood.[17] Shooting data include all reported shooting incidents, most of which do not involve police officers, and contain information on whether the shooting results in a civilian's death, which

[14] These proprietary data also contain the exact location of where the stop occurred, while publicly-released stop data code stop locations at the nearest intersection.

[15] An audit of stops conducted from 2016-2019 found that officers did not report up to 30% of stops conducted. For this reason, we analyze stop outcomes data prior to 2016. The audit, however, indirectly confirmed that stop rates had indeed fallen by over 95%, since at the time of the audit stops were roughly 3% of pre-reform levels.

[16] Unlike crime reports, however, arrest records have complete race information, allowing us to explore how the racial composition of offenders evolves throughout the study period.

[17] We can, however, examine trends in approximate clearance rates at the city-level by dividing the total number of arrests by the total number of reported crimes over a given period (see Figure **??**).

hereafter we refer to as "killings".[18] Finally, we supplement the above data with demographic information about the residential population of each tract using the 2010 Census, the 2010 American Community Survey five-year estimates, and NYC Housing Authority maps.

Crime Outcomes

We first examine the impact of stop rates on two broad crime categories – felonies and violent misdemeanors, and, non-violent misdemeanors and violations. These categories are associated with substantially different penalties and social costs. In New York City, the most severe misdemeanor crimes carry a weight of up to 364 days in jail and a fine of up to \$1000.[19] Violations are even less severe, carrying a weight of up to 15 days of jail time. Agan et al. (2021b) show that non-prosecution of marginal non-violent misdemeanors leads to large reductions in the likelihood that a defendant is charged with a new criminal complaint over the subsequent two years. These findings suggest that charges for some low-level offenses, such as those associated with graffiti or marijuana possession, may even impose negative costs on society if future behavioral responses are appropriately incorporated.

Felonies and violent misdemeanors, on the other hand, involve serious crimes that impose substantial costs on society. For example, Palmer et al. (2017) estimate that the direct costs of the average violent felony crime is \$67,794. We additionally report impacts on two mutually-exclusive sub-categories – "major felonies" and "non-major felonies and violent misdemeanors" since the police often use major felony crime to track progress in a given area.[20] Major felonies include murder and non-negligent manslaughter, rape, robbery, felony assault, burglary, grand larceny, and grand larceny of motor vehicles.[21]

B High School Enrollment Data

We employ administrative enrollment records from the NYC Department of Education to examine the impact of changes in stop rates on high school drop out rates. We focus on high school students since stop rates diverge at age 14 and are the highest at ages 18-20.[22] Data were obtained through NYU's Research Alliance for New York City Schools and contain detailed enrollment information on the universe of high school students enrolled in

[18]The NYPD did not start using ShotSpotter technology until late 2015, and rolled out this technology broadly in higher-crime areas in 2016. Thus, most shootings during our study period were reported through traditional channels.

[19]In New York state, misdemeanor theft involves goods worth no more than \$1,000.

[20]For example, commander performance was often judged at CompStat meetings by changes in major felonies within their designated area (e.g. precinct or borough).

[21]These seven offenses roughly track the sum of violent and property crime measures in the FBI's Uniform Crime Reports, making effects on major felonies readily comparable to other studies examining crime deterrence effects (e.g. Chalfin and McCrary (2018)).

[22]See Figure A.2 for race-specific age profiles of stop rates.

NYC public schools from 2001 through 2019. Enrollment status is given for each student in October and June of every school year. As required by the city, school administrators must note the specific reason why a student is not enrolled in a given semester. We use these enrollment categories to construct our two main outcomes.

The first outcome is an indicator of whether a student was discharged from high school based on an institutional directive by a non-DOE agency, court, or other authority (hereafter, "discharged by institutional directive"). This includes students who are in the custody of the Department of Corrections, or who attend special programming under the supervision of the NYS Department of Mental Health or Office of Children and Family Services. It is important to note that this outcome does not include situations where a student is arrested after they stop attending school and therefore likely underestimates the direct effect of stop rates on the likelihood of dropping out because of an arrest.[23]

The second outcome indicates whether the student dropped out of high school or was discharged by institutional directive. We set each outcome indicator to 1 if and only if the student switches from being listed as "enrolled" in October of a given school year to being marked as discharged by October of the subsequent school year. This definition allows us to isolate the exact timing of when students stop attending high school.[24] Across both outcomes, less than 7% (18%) of students go on to graduate high school within four (six) years. These data also contain detailed demographic information, including student race, sex, month of birth, languages spoken at home, and the Census tract of the student's home address.

We restrict the sample to high schools that were operational from school years 2010 through 2017, which includes 505 schools and 90.2% of all students enrolled in the public school system over this period.[25] We further restrict the sample to students who are enrolled in October of their initial 9th grade school year, and keep the first four school years of their high school tenure.[26] This leaves a final sample of 2,092,366 student-year observations that span eight school years.

[23] Students are listed as having "dropped out" after 20 days of consecutive absence.

[24] Note that students may drop out more than once if they re-enroll and dropout again.

[25] Throughout this section we refer to school years by spring-term year. For example, the 2010 school year refers to the 2009-2010 school year.

[26] This includes observations for students who have left the NYC school system. For example, a student who begins 9th grade in 2012 will be observed four times in our data, spanning 2012 through 2015 school years, even if they choose to leave school after their junior year.

IV Crime Deterrence Effects

A Empirical Strategy

To estimate the impact of stops on neighborhood crime, we implement a flexible difference-in-differences (DD) framework that compares changes in crime rates induced by *Floyd* in neighborhoods that had *similar crime rates but substantially different stop rates prior to the reform*. This approach exploits the fact that a significant portion of the variation in neighborhood stop rates pre-*Floyd* cannot be explained by observable crime rates. Prior to the reform, "treatment" neighborhoods have high stop rates conditional on crime (H), while our "control" neighborhoods exhibit medium stop rates conditional on crime (M). After *Floyd*, stop rates in both groups fall to very low levels (L). Therefore, we estimate the crime deterrence effect of conducting $H-M$ additional stops prior to the reform by subtracting changes observed in control neighborhoods from changes observed in treated neighborhoods

$$\left[\frac{d\text{Crime}}{ds_n}\right]_{H\to M} = \left[\frac{d\text{Crime}}{ds_n}\right]_{H\to L} - \left[\frac{d\text{Crime}}{ds_n}\right]_{M\to L} \tag{1}$$

where we expect $\left[\frac{d\text{Crime}}{ds_n}\right]_{H\to M} > 0$ if the additional stops have a crime deterrence effect.

To separate neighborhoods into treatment and control groups, we first split the sample into three distinct periods. The first period is the "training" period and spans January 1st, 2006 through April 15th, 2008. The second period is the "pre-period" and contains the four years leading up to the reform. The third and final period is the post-*Floyd* period and includes the five years after the reform (April 16th, 2012 through April 15th, 2017).

Restricting data to the training period, we identify neighborhoods with disproportionately high stop rates conditional on observed crime rates by running the following least-squares regression for each tract n in week-year t:

$$Stops_{n,t} = \alpha_S Shootings_{n,t} + \alpha_F Felonies_{n,t} + \alpha_M Misdemeanors_{n,t} + \Gamma X_n' + \delta_t + \epsilon_{n,t} \tag{2}$$

$Shootings_{n,t}$, $Felonies_{n,t}$, and $Misdemeanors_{n,t}$ are vectors that include up to third-order polynomial terms as well as additional linear measures of average crime rates of Census tracts within 0.5 and 1.0 miles of a given neighborhood.[27] δ_t are week-year fixed effects and X_n' includes neighborhood deciles of total land area and total population, as recorded in the 2010 Census. For each n, we then calculate the mean stop residual during the training period – $\bar{\epsilon}_n^{Train} = \frac{\sum_{t\in Train}\epsilon_{nt}}{\sum_{t\in Train} I(t)}$ – and split neighborhoods into "treatment" and "control" groups at

[27] Specifically, we calculate average crime rates in Census tracts with centroids within 0.5 and 1.0 miles from the centriod of a given Census tract n.

Table II: Balance between Treatment and Control Neighborhoods prior to *Floyd*

	Treatment	Control
Panel A: Stop, Question, & Frisk		
Stops	98.680	50.037
Frisks and Searches	55.389	30.151
Uses of Force	21.282	13.351
% Stops with Arrest or Weapons/Drugs Found	0.101	0.117
% Stops with Arrest	0.068	0.081
% Stops with Weapon Found	0.013	0.015
% Stops with Durgs Found	0.020	0.021
Mean Stops per Officer in 30 Days	1.518	1.394
Number of Officers with Stop Made in 30 Days	17.762	9.730
Panel B: Reported Crimes		
Shootings	0.253	0.217
Major Felonies	11.369	13.921
Other Felonies	4.837	5.133
Misdemeanors & Violations	38.921	45.254
All Black Arrests / All White Arrests	2.176	1.614
All Hispanic Arrests / All White Arrests	1.512	1.466
Panel C: Neighborhood Characteristics		
Census Tract Area in Square Miles	0.169	0.106
Population in 2010 Census	3391.7	4172.9
% Black	0.268	0.223
% Hispanic	0.268	0.265
% Census Tract Area Belonging to NYC Housing Authority	0.037	0.005
Median Househould Income	57030.5	57110.8
% Residents Living Below Federal Poverty Line	0.161	0.158
% Residents with Bachelor's Degree or Higher	0.320	0.362
% Residents with Less than HS Degree	0.182	0.172
% Residents with SSI	0.070	0.068
% Residents Living in Single-parent Household	0.315	0.291

Notes: This table reports neighborhood means for Treatment and Control neighborhoods during the three years prior to *Floyd*. *Panel A* reports SQF outcomes, *Panel B* provides means for various reported crimes and racial arrest ratios, while *Panel C* provides means of neighborhood characteristics. N = 2,058 Census tracts split evenly into 1,079 Treatment and 1,079 Control neighborhoods.

the median. Our findings are robust to alternative choices of covariates in Equation 2 and to using larger geographic units, such as precinct-level variation in stop exposure.

Table II shows that relative to control neighborhoods, treatment neighborhoods experience almost twice as many stops, 80% more frisks, and 60% more uses of force during the pre-period. On average, stops are also 14% less likely to detect criminal behavior in treated neighborhoods. Differences in officer presence explain much of the difference in stop rates, as the number of officers conducting stops in a 30 day-span is 80% higher in treated neighborhoods, while the average number of stops per officer is only 9% higher in treatment neighborhoods. Reported crime rates are comparable across study groups in the pre-period. While treatment neighborhoods experience slightly more weekly shootings per square mile, all other crime categories (major felonies, other felonies, and misdemeanors and violations per 1mi^2/week) are slightly higher in the control group. A variety of neighborhood characteristics are also comparable across study groups, including median household income and the percent of residents who are Hispanic, live in single-parent households, or live below the federal poverty line. Treated neighborhoods differ in that residents are more likely to be Black, a larger fraction of the area contains NYC Housing Authority developments, and are less densely populated.

With treatment defined, we restrict the data to the four years before the reform through five years after the reform. We then compare how outcomes evolve for treatment neighborhoods relative to control neighborhoods by implementing the following difference-in-differences regression:

$$Y_{nt} = \sum_{\tau \neq -1} \beta_\tau TREAT_\tau + \delta_n + \delta_{p \text{ x } t} + \epsilon_{nt} \tag{3}$$

where τ denotes event-time in years relative to the onset of *Floyd*. δ_n are neighborhood fixed effects and $\delta_{p \text{ x } t}$ are precinct-by-week-year fixed effects. A neighborhood is defined as a Census tract, which covers an average area of 0.12 square miles or four-by-four blocks. $TREAT_\tau$ are relative time to treatment indicators that equal 1 for treatment neighborhoods in year τ. The coefficients of interest, $\{\beta_\tau\}$, then estimate the average change in Y between year τ and the reference year in treated neighborhoods relative to that same change over time in control neighborhoods in the same precinct p. We set the reference year as two years prior to *Floyd* (e.g. $\tau = -1$) since the mean stop rate in this year is close the four-year pre-period mean. All standard errors are clustered at the precinct-level, allowing for serial correlation within each of the 76 precincts (Bertrand et al., 2004). Our *identifying assumption* is therefore the standard common trends assumption: treatment and control neighborhoods would exhibit common trends in crime outcomes absent *Floyd*. With this assumption, any differential changes in crime rates in response to *Floyd* can then be interpreted as the crime that was deterred by higher stop rates during the pre-*Floyd* period.

B Main Results

We first explore the causal effect of the *Floyd* by estimating Equation 3 on stops, frisks, felonies and violent misdemeanors, and non-violent misdemeanors and violations. All data are at the tract-by-week level and outcomes are normalized to weekly rates per square mile. The omitted year is two years prior to *Floyd*. Estimates are displayed in Figure IV.

Prior to *Floyd*, Panels A and B display small positive trends in stops and frisks for treated neighborhoods relative to control neighborhoods, suggesting that our approach slightly understates treatment-control differences in stop rates. Due to our choice of $\tau = -1$ as the reference period, however, pre-treatment coefficients are jointly insignificant for both stops ($F = 0.06, p = 0.814$)) and frisks ($F = 0.31, p = 0.580$). The first dashed line represents the police's initial response to the court allowing expert testimony. In the following year, weekly stop (frisk) rates fell by an additional 20 stops (10 frisks) per square mile in treated neighborhoods relative to control neighborhoods. The second dashed line marks the post-ruling phase. By year three, weekly stop (frisk) rates have fallen by an additional 50 stops (28 frisks) in treated neighborhoods relative to control neighborhoods – a difference that is roughly the magnitude of the control group's pre-period mean.

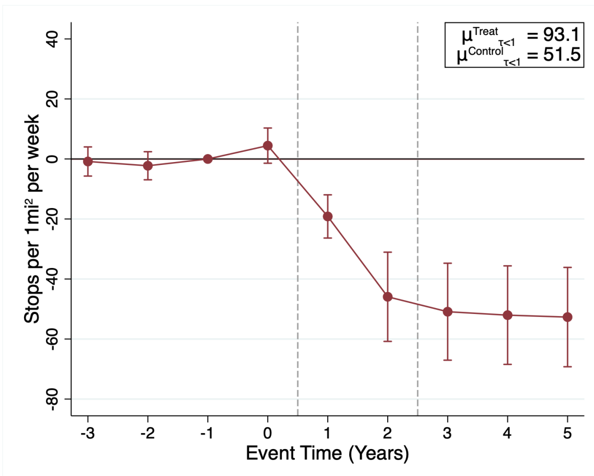
Panel C displays impacts on our primary crime outcome – reported felonies and violent misdemeanors. For $\tau < 1$, there is no evidence of differential pre-trends. All treatment coefficients are less than 0.24 weekly crimes per square mile and each coefficient is not statistically significant, even at the 30-percent level.[28] Pre-treatment estimates are also jointly insignificant ($F = 0.00, p = 0.996$). Panel D shows a similar lack of pre-trend differences for reported non-violent misdemeanors and violations, as pre-treatment point estimates alternate signs and are jointly insignificant ($F = 0.78, p = 0.381$). We find no evidence of increases in felonies or violent misdemeanors during the post-*Floyd* period. Post-treatment coefficients alternate signs, are statistically insignificant, and are close to zero.

Table III reports coefficients from Equation 3 using a single post-*Floyd* treatment indicator. We estimate an effect of -0.097 crimes per 1mi^2 or -0.4% of the pre-period mean ($p = 0.672$). Put differently, treated neighborhoods that had twice as many stops in the four years prior to *Floyd* do not experience relative increases in serious crimes. In fact, our 95% confidence interval rules out a 1.5% increase. To put these effects in context, New York City saw a 38% decline in felony crime rates from 2000 to 2010. These estimates suggest that increasing stop rates did not play a significant role in reducing serious crime.

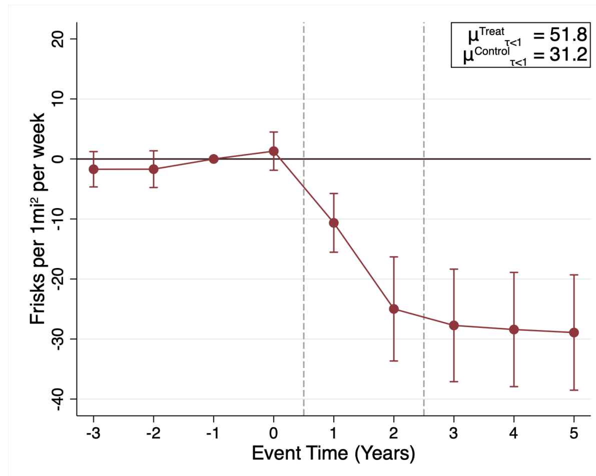
[28] The largest coefficient is just $\frac{1}{95}$th of the pre-treatment mean.

Figure IV: *Floyd* Difference-in-differences Estimates

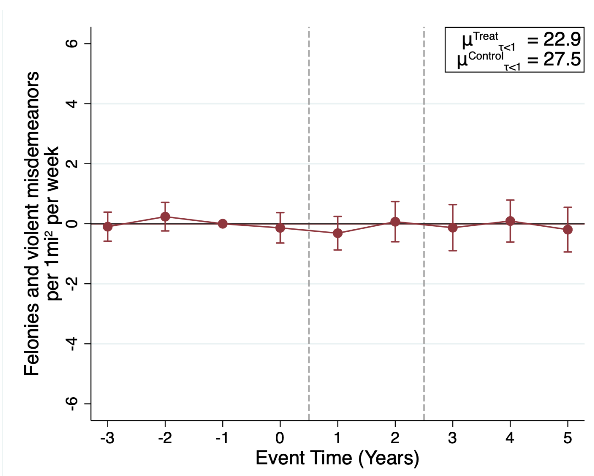
(a) Stops



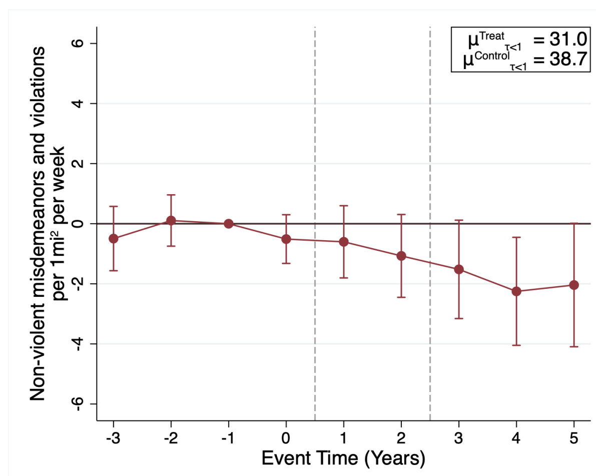
(b) Frisks



(c) Felonies & Violent Misdemeanors



(d) Non-violent Misd. & Violations



Notes: This figure graphs coefficients from Equation 3 on stops, frisks, felonies and violent misdemeanors, and non-violent misdemeanors and violations. Maroon dots denote point estimates and whiskers show 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the precinct level. Pre-period outcome means are given for treatment and control neighborhoods in the northeast corner of each figure.

Table III: Effect of *Floyd* on Neighborhood Crime

	Pre-period Mean		$\beta_{\tau \geq 1}$
	Control	Treatment	
	(1)	(2)	(3)
Stops per 1mi^2/week	51.471	93.143	-44.475***
			(7.031)
Frisks per 1mi^2/week	31.196	51.793	-23.611***
			(3.964)
Uses of force per 1mi^2/week	13.837	20.302	-10.490***
			(2.215)
Stop-related arrests per 1mi^2/week	3.429	5.366	-2.096***
			(0.348)
Stop-related court summonses per 1mi^2/week	3.254	5.844	-2.660***
			(0.463)
Shootings per 1mi^2/week	0.235	0.240	0.007
			(0.014)
Killings per 1mi^2/week	0.045	0.043	0.000
			(0.004)
Felonies & violent misd per 1mi^2/week	27.457	22.938	-0.097
			(0.229)
Major felonies per 1mi^2/week	14.406	11.298	0.123
			(0.144)
Non-major fel & violent misd per 1mi^2/week	13.052	11.640	-0.220
			(0.161)
Non-violent misd & violations per 1mi^2/week	38.674	31.031	-1.272*
			(0.669)

Notes: This table reports estimates for various SQF and reported crime outcomes from Equation 3, except that we collapse the five-year post-*Floyd* period into a single indicator. Columns (1) and (2) report outcome means during the four years prior to *Floyd* for control and treatment neighborhoods, respectively. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 2,058 Census tracts with 50-50 % split of Treatment and Control neighborhoods.

Also contrary to the crime deterrence hypothesis, Panel D reveals that treated neighborhoods experience a relative reduction in non-violent misdemeanors and violations following the reform. We show in Table III that these reductions can be entirely explained by reductions in stop-related arrests, which predominantly involve non-violent misdemeanor offenses.[29] These findings suggest that the main result of heightened stop rates is increased detection and enforcement of minor offenses.

A full summary of our findings is provided in Table III. In addition to our main outcomes, we report impacts on frisks, uses of force, shootings, and killings, and disaggregate reported crimes into major felonies, non-major felonies and violent misdemeanors, and non-violent misdemeanors and violations. Differences in the reduction of stops, frisks, and uses of force between treated and control neighborhoods are about double the size of pre-period outcome levels in the control group. Estimates of effects on shootings, killings, major felonies, and non-major felonies and violent misdemeanors confirm the notion that serious crime was unaffected by *Floyd*; point estimates are close to zero and statistically insignificant.

One may be concerned that aggregate differences across treatment and control groups may mask crime increases in higher-crime neighborhoods. We directly investigate this question by estimating effects separately for the highest felony-crime quartiles in Table A.1 (where quartiles are measured using training period data). The accompanying figure – Figure A.3 – displays coefficients from Equation 3 for neighborhoods with above-median felony crime rates during the training period. There is no evidence of differential increases in felonies and violent misdemeanors, shootings, killings, major felonies, or non-major felonies and violent misdemeanors in higher-crime neighborhoods. Additionally, aggregate *reductions* in non-violent misdemeanors and violations are driven almost entirely by a 7% reduction in crimes in the highest-crime quartile ($p = 0.01$).

Crime Reporting

While it is possible that null effects reflect offsetting changes in civilian crime reporting and criminal behavior, this seems unlikely in our setting. A reduction in aggressive interactions with the police is likely to raise crime reporting rather than reduce reporting tendencies. Increased reporting in the post-period would serve to inflate our estimates of crime deterrence, as treated neighborhoods observe a relatively larger reduction in aggressive police interactions. We also test crime reporting directly by estimating the reform's effect on outcomes that are less likely to go unreported. These include shootings, murders, and thefts of expensive goods, such as cars or goods worth over $1,000. Table IV shows that all outcome coefficients are close to zero and statistically insignificant.

[29] Stop-related court summonses are not included in our crime reports data, while stop-related arrests are.

Table IV: Effect of *Floyd* on Crime Categories Less Affected by Reporting Concerns

	Pre-period Mean		$\beta_{\tau \geq 1}$	P-value
	Control	Treatment		
	(1)	(2)	(3)	(4)
Killings per 1mi^2/week	0.045	0.043	0.000	0.970
			(0.004)	
All murders per 1mi^2/week	0.058	0.056	-0.006	0.168
			(0.004)	
Grand larceny crimes per 1mi^2/week	6.751	4.839	0.046	0.685
			(0.113)	

Notes: This table reports difference-in-differences estimates on various categories of crimes per square mile using Equation 3 with a singular post-*Floyd* indicator. Killings refer to deadly shooting incidents, murders refer to all murder and non-negligent manslaughter crimes. Grand larceny includes all car thefts as well as thefts of property above $1,000 in value. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness

Table A.2 displays robustness to a variety of alternative specifications, treatment definitions, and sub-samples. Column 1 reports estimates from our preferred specification described in Equation 3 using a singular post-treatment indicator variable. Column 2 and 3 replace precinct-by-week-year fixed effects with week-year fixed effects and borough-by-week-year fixed effects, respectively. To confirm that our results are not sensitive to alternative ways of estimating pre-*Floyd* stop residuals, Columns 4 through 6 explore alternative treatment definitions. Column 4 assigns treatment based on precinct-level variation in stop rates during the training period by running Equation 2 at the precinct-level.[30] Columns 5 and 6 explore alternative ways to define treatment at the tract-level. Column 5 removes all covariates from Equation 2 except for mean shootings, major felonies, non-major felonies, and misdemeanors per 1mi^2/week and week-year fixed effects. Column 6, on the other hand, adds various time-invariant neighborhood characteristics to Equation 2.[31] Finally, Columns 7 to 9 restrict the sample to higher-crime neighborhoods where stops are most likely to deter crime. Columns 7, 8, and 9 respectively restrict the sample to neighborhoods that are ranked in the top 75%, 50%, and 25% of neighborhoods in terms of mean felony crimes per

[30] Given that treatment variation then occurs at the precinct-level, we replace precinct-time fixed effects with borough-time fixed effects.

[31] These include indicators for having 0-10% or $\geq$ 10% of tract area covered by public housing, and decile-fixed effects for the following neighborhood characteristics: % Black, % Hispanic, % families living below the poverty line, median household income, and % of adult residents with less than a HS degree.

1mi^2/week during the training period.

Across all specifications, reductions in stop rates are substantially larger in treated neighborhoods. The average additional decline in treated relative to control neighborhoods ranges from 46 - 162% of the pre-period control group mean. Panel C displays point estimates of effects on major felonies. Importantly, these coefficients are statistically insignificant and close to zero, spanning +1.0 to -1.3 percent of the treatment group's pre-period mean. In fact, our preferred specification is the largest positive estimate among the group. These findings support the notion that major felonies did not differentially increase in tracts (or precincts) that experienced significantly larger declines in stop rates. Across seven of the nine specifications presented in Panel B, felonies and violent misdemeanors are similar across treatment and control groups. Columns 5 and 6 report modest reductions in felonies and violent misdemeanors of about 2% of the pre-period mean, driven by declines in non-major felonies and violent misdemeanors.

Figure A.4 displays a sensitivity analysis per Roth and Rambachan (2021) that tests the robustness of effects on felonies and violent misdemeanors to parallel trend violations. This figure reports 95% confidence intervals allowing for a change in linear slope between any two consecutive periods of up to M. Given the lack of pre-trends, we find the confidence sets for each post-reform coefficient is centered around zero. Setting M at 0.20 crimes per 1mi^2/week, the adjusted 95% confidence interval on the first-year coefficient rules out increases greater than 1 crime per 1mi^2/week or 2.3% of the pre-period treatment mean. These confidence intervals gradually grow with each additional year since the reform; year three estimates, for example, can rule out increases greater than 3 crimes per 1mi^2/week (6.9%). The precision of four- and five-year estimates requires stronger assumptions on trend violations, closer to the standard assumption of parallel trends.

C Police Surges

Does an increase in pedestrian stops *and officer presence* deter neighborhood crime? The lack of crime responses to the *Floyd* reform does not provide information on whether street patrols, more generally, are ineffective at deterring crime. Following the reform, patrol officers continued to be assigned to crime hot-spots even though they stopped pedestrians at reduced rates in these neighborhoods. To decouple the crime effect of stop rates from officer presence, we assess the impact of police surges prior to the reform that simultaneously increased officer assignments and stop rates in higher-crime neighborhoods. As discussed earlier, neighborhoods selected as Impact Zones received an influx of officers that increased foot patrols, conducted a high volume of pedestrian stops, and strictly enforced minor infractions. Table V shows that, prior to assignment, neighborhoods selected as "Impact Zones" have higher rates of crime and stops, and are more disadvantaged along a variety of measurable characteristics (i.e. median household income, federal poverty rate, education).

We begin by estimating crime deterrence effects using a standard difference-in-differences framework:

$$Y_{n,t} = \sum_{\tau \neq 0} \beta_\tau IZ_\tau + \Gamma X'_{nt} + \delta_n + \delta_{p \text{ x } t} + \epsilon_{nt} \tag{4}$$

where τ denotes six-month periods relative to neighborhood n's next Impact Zone assignment, and IZ_τ indicates whether neighborhood n has more than 5% of its total area covered by an Impact Zone in the post-assignment window ($\tau \geq 1$). $\delta_{p \text{ x } t}$ are precinct-by-week-year fixed effects and δ_n are neighborhood fixed effects. We restrict our attention to the three years surrounding a given neighborhood's assignment event (i.e. $\tau \in [-5, 6]$). X'_{nt} is a control vector that includes indicators equal to one for periods outside of this three-year window, fixing the reference period as the six months prior to assignment (i.e., $\tau = 0$). Standard errors are clustered at the precinct-level (Bertrand et al., 2004).

Our preferred specification adapts Equation 4 by dropping pre-treatment indicators ($\sum_{\tau=-5}^{-1} \beta_\tau IZ_\tau$) and adding a term that allows Impact Zone neighborhoods to exhibit linear differences in pre-trends:

$$Y_{n,t} = \sum_{\tau=1}^{6} \beta_\tau IZ_\tau + \beta_{trend}[IZ(-5 \geq \tau \leq 6) \text{ x } \tau] + \Gamma X'_{nt} + \delta_n + \delta_{p \text{ x } t} + \epsilon_{nt} \tag{5}$$

where $[IZ(-5 \geq \tau \leq 6) \text{ x } \tau]$ captures the linear difference in slopes for treated neighborhoods relative to control neighborhoods during the three-year pre-period. We take this approach since Figure V reveals felony and violent misdemeanor rates are increasing at a faster rate in Impact Zone neighborhoods relative to other control neighborhoods, as neighborhoods were selected as Impact Zones due to a rise in criminal activity. Our coefficients of interest (β_τ) then represent the mean differential change in Y_{nt} between event-time τ relative to a linear trend observed during the pre-period among treated neighborhoods relative to that same change in Y_{nt} over time among untreated neighborhoods in precinct p. Here, the *identifying assumption* is that differences in neighborhood outcomes between treatment and control neighborhoods would have continued to follow a linear trend absent Impact Zone assignment. We understand that reversion to the mean may drive these results to be upward biased, but are less concerned about this issue since we observe a relatively flat pre-trend for more severe crimes (i.e., major felonies).

Figure V graphs coefficients from Equation 4 on various policing and crime measures. Panels A through C illustrate a strong first stage. For $\tau < 0$, there are no visible pre-trend differences in area covered by an Impact Zone, the number of officers conducting stops, or

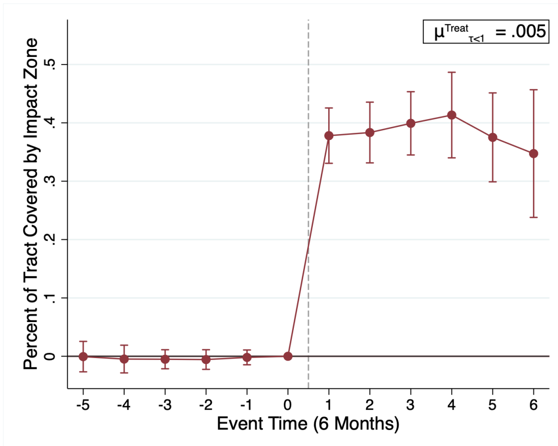
Table V: Stops, Crimes, and Characteristics of *Impact Zone* Neighborhoods

	Not Impact Zone			Impact Zone	
	> 1.0 Mi	1.0 to 0.25 Mi	0.25 to 0.0 Mi	$-5 \leq \tau \leq 0$	$1 \leq \tau \leq 6$
Panel A: Stop, Question, & Frisk					
Stops	34.709	76.267	173.217	155.975	220.969
Frisks and Searches	17.739	43.089	102.157	87.322	126.627
Uses of Force	7.130	17.630	49.300	38.217	51.036
% Stops with Arrest or Weap/Drugs Found	0.105	0.107	0.086	0.087	0.072
% Stops with Arrest	0.073	0.074	0.058	0.057	0.047
% Stops with Weapon Found	0.011	0.013	0.011	0.012	0.012
% Stops with Drugs Found	0.021	0.021	0.017	0.018	0.014
Mean Stops per Officer in 30 Days	1.388	1.470	1.521	1.529	1.571
Number of Officers Making Stops in 30 Days	8.079	13.902	26.190	24.951	33.664
Panel B: Reported Crimes					
Murders	0.019	0.058	0.137	0.139	0.132
Shootings	0.094	0.316	0.656	0.706	0.769
Major Felonies	9.317	17.766	25.357	22.219	22.226
Other Felonies	3.124	6.386	12.372	11.452	11.802
Misdemeanors & Violations	27.286	52.741	94.133	79.996	84.643
Arrest Ratio: Black / White	1.015	2.495	4.703	5.078	5.786
Arrest Ratio: Hispanic / White	0.875	2.124	3.574	4.034	3.602
Panel C: Neighborhood Characteristics					
Census Tract Area in Square Miles	0.169	0.092	0.067	0.072	0.076
Population in 2010 Census	3600.0	3912.0	4384.7	4327.3	4250.5
% Black	0.172	0.298	0.381	0.474	0.517
% Hispanic	0.211	0.337	0.445	0.394	0.351
% of residences in NYCHA Housing	0.015	0.022	0.036	0.043	0.045
Median Household Income	63644.2	53639.3	38119.4	35914.7	35900.5
% Living Below Fed Pov Line	0.123	0.183	0.258	0.274	0.277
% with Bach Degree or Higher	0.380	0.337	0.217	0.197	0.188
% without HS Degree	0.148	0.198	0.269	0.258	0.247
% with SSI	0.060	0.072	0.099	0.107	0.096
% Single-parent Households	0.243	0.346	0.447	0.485	0.488
N (Tract x Week-years)	463,527	146,368	34,929	50,699	36,039

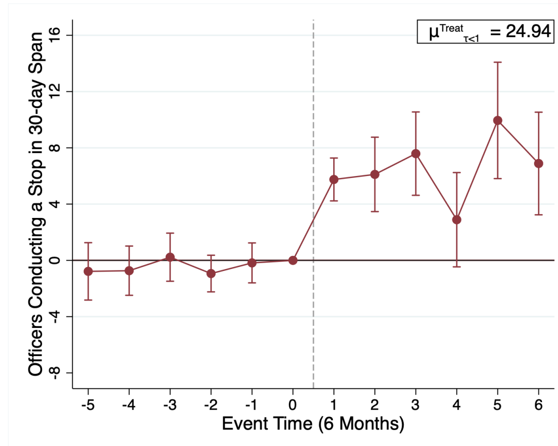
Notes: This table reports means of outcomes and neighborhood characteristics for neighborhoods by proximity to Impact Zones and time relative to Impact Zone assignment. Data are at the tract-week-year level and span Week 2 of 2006 through Week 28 of 2012. Column (1) includes Census tracts more than 1.0 miles away from an Impact zone; Column (2) includes tracts within 1.0 to 0.25 miles of an Impact Zone; Column (3) includes tracts within 0.25 to 0.0 miles of an Impact Zone; Column (4) includes week-year observations of Census tracts during the three years ($-5 \geq \tau \leq 0$) prior to Impact Zone assignment; Column (5) includes weeks in the three years following assignment ($1 \geq \tau \leq 6$). *Panel A* reports SQF outcomes, *Panel B* provides means for various reported crimes and racial arrest ratios, while *Panel C* provides means of various neighborhood characteristics. The final row reports the number of observations at the tract-by-week-year level for each study group.

Figure V: Standard Difference-in-Differences *Impact Zone* Estimates

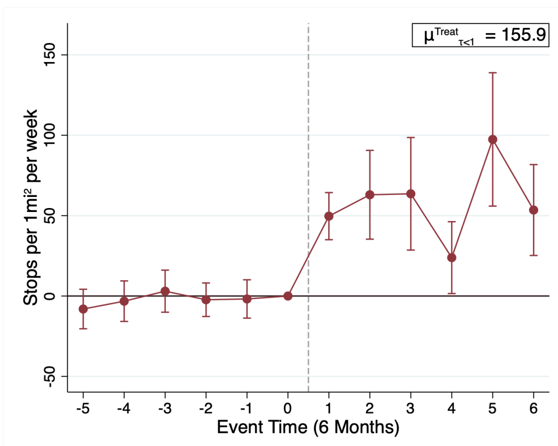
(a) Fraction of CT Covered by an Impact Zone



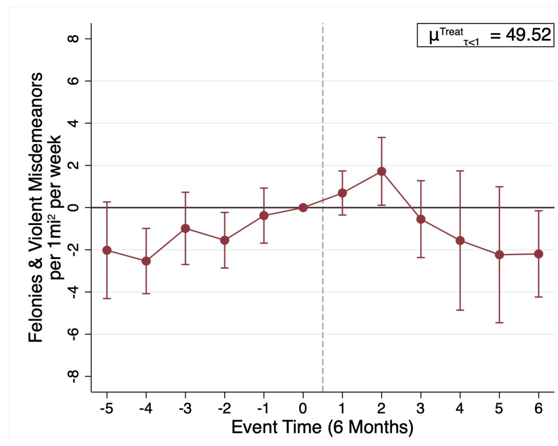
(b) Officers Conducting a Stop in 30-day Span



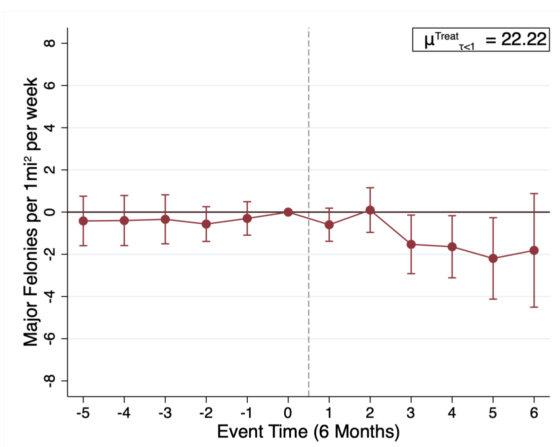
(c) Stops



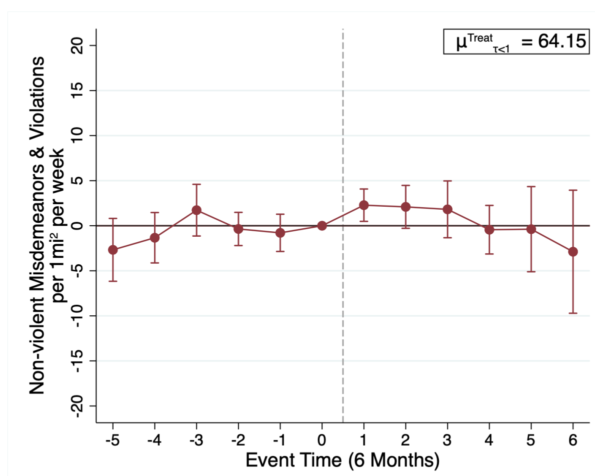
(d) Felonies & Violent Misdemeanors



(e) Major Felonies



(f) Non-violent Misd. & Violations



Notes: This figure graphs coefficients from a Equation 4. Maroon dots denote point estimates and whiskers show 95% confidence intervals. The dashed line denotes the timing of Impact Zone assignment. All standard errors are clustered at the precinct level. *Impact Zones* refer to Census tracts with > 5% of their area covered by an Impact Zone. Pre-period outcome means are given for treatment neighborhoods.

the number of weekly stops per square-mile. Following assignment, the average "treated" neighborhood has 40% of it's area covered by an Impact Zone, and both the number of officers conducting stops and weekly stop rates jump up by about one-third.

Panels D through F respectively display impacts on reported felonies and violent misdemeanors, major felonies, and non-violent misdemeanors and violations. While there are no visible pre-trend differences between study groups for major felonies or non-violent misdemeanors and violations, there is a modest positive pre-trend difference in felonies and violent misdemeanors. This likely stems from the fact that Impact Zones were, in part, selected because of abnormal increases in crime rates.

Table VI reports coefficients from Equation 5 across a variety of policing and crime outcomes. We aggregate treatment coefficients into post-assignment year indicators in order to improve statistical power. Sharp increases in stop rates are primarily driven by increases in the number of patrol officers conducting stops per month, which rose by 25-38% after assignment, compared to only a 4-6% increase in mean number of monthly stops conducted per officer. Frisks and court summonses also increase by more than one-third following assignment, while stop-related arrest rates are unaffected, implying that additional stops did not translate into additional arrests.

One year after assignment, there is no detectable impact on felonies and violent misdemeanors. However, this masks a small negative but statistically insignificant impact on major felonies and a positive and significant impact on non-major felonies and violent misdemeanors ($p < 0.01$). The latter is potentially the result of increased detection and/or stricter enforcement. Treatment effects become negative for both measures in year two, and by year three, major felonies per 1mi^2/week fall by 11% and non-major felonies and violent misdemeanors fall by 8%. Over the entire post-assignment period, reductions in felonies and violent misdemeanors are jointly significant ($p = 0.011$), with the largest reductions observed for major felonies ($p = 0.015$). As we found in the *Floyd* analysis, coefficients on non-violent misdemeanors and violations are positive in the first year post-assignment, potentially due to increased *detection*, but flip signs by year three. Post-assignment coefficients on shootings are positive but are jointly insignificant ($p = 0.236$) and coefficients on killings are close to zero and jointly insignificant.

In contrast to our *Floyd* estimates, we find that simultaneously increasing stop rates and officer presence in higher-crime neighborhoods effectively deters major felony crime. For each additional patrol officer conducting a positive number of stops per month, 21.65 major felonies are deterred each year per 10,000 neighborhood residents. This estimate is strikingly similar to the macro elasticity estimated by Mello (2018), which compares cities just beyond the COPS grant application cutoff to those just below. Mello estimates that each additional

Table VI: Effect of Impact Zone Assignment on Neighborhood Policing and Crime

	Pre-period Mean	Year 1	Year 2	Year 3	P-value
Policing	(1)	(2)	(3)	(4)	(5)
Fraction of Tract Covered by Impact Zone	0.005	0.379***	0.402***	0.361***	0.000
		(0.024)	(0.030)	(0.044)	
Stops per $1mi^2$/week	155.943	55.470***	50.402***	86.492***	0.000
		(9.102)	(14.953)	(19.766)	
Officers with a Stop in 30-day Span	24.947	6.133***	6.099***	9.411***	0.000
		(0.855)	(1.490)	(2.340)	
Mean Stops per Officer	1.529	0.084**	0.061**	0.086**	0.000
		(0.018)	(0.026)	(0.030)	
Frisks per $1mi^2$/week	87.305	30.730***	29.456***	31.766***	0.000
		(6.205)	(8.863)	(10.937)	
Uses of Force per $1mi^2$/week	38.210	12.700***	6.990	5.020	0.046
		(4.227)	(4.867)	(7.321)	
Stop-related Arrests per $1mi^2$/week	7.985	0.696	-0.536	-1.276	0.555
		(0.513)	(0.874)	(0.890)	
Stop-related Court Summonses per $1mi^2$/week	11.335	3.731***	5.930***	4.311**	0.001
		(0.898)	(2.053)	(1.659)	
Crime					
Felonies & Violent Misd per $1mi^2$/week	49.520	0.550	-2.287**	-4.359***	0.011
		(0.488)	(1.099)	(1.244)	
Major Felonies per $1mi^2$/week	22.220	-0.381	-1.764**	-2.422**	0.015
		(0.356)	(0.739)	(1.097)	
Non-major Fel & Violent Misd per $1mi^2$/week	27.301	0.931***	-0.523	-1.937**	0.346
		(0.286)	(0.750)	(0.894)	
Non-violent Misd & Violations per $1mi^2$/week	64.153	2.333***	0.953	-1.319	0.656
		(0.940)	(1.533)	(2.523)	
Shootings per $1mi^2$/week	0.706	0.172**	0.169	0.154	0.236
		(0.085)	(0.137)	(0.219)	
Killings per $1mi^2$/week	0.139	-0.022	0.011	-0.024	0.820
		(0.033)	(0.056)	(0.080)	
Observations (tract-weeks)	86,759				
Census Tracts	385				

Notes: This table reports coefficients from Equation 5 for various policing and crime measures. We collapse post-assignment six-month treatment indicators into three indicators denoting each year post-assignment. "Pre-period mean" refers to the outcome mean in treated neighborhoods in the three years leading up to assignment. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Treatment is defined as $> 5\%$ of Census tract area covered by an Impact Zone. "P-val" refers to the p-value from an F-test of the null hypothesis that post-treatment indicators are jointly zero.

sworn officer reduces major felony crimes by 19.66 crimes per 10,000 residents.[32]

Robustness

The main concern with controlling for linear pre-trends is regression to the mean. Since treated neighborhoods are, in part, chosen because they display abnormally high felony crime rates in the months leading up to assignment, counterfactual crime rates might have naturally subsided in the post-period regardless of Impact Zone assignment. However, we do not observe large positive pre-trends for major felony crimes, and post-reform coefficients are significantly lower than the lowest pre-trend coefficient (see Figure V). Regression to the mean cannot explain the magnitude of reductions observed for felonies and violent misdemeanors. We show this formally by allowing for parallel trend violations a la Roth and Rambachan (2021).[33] Figure A.5 shows that second- and third-year coefficients are significant at the 5% level if we allow for differential pre-trend slopes of +/- 0.35 crimes per 1mi^2/week and +/- 0.45 crimes per 1mi^2/week, respectively. These pre-trend violations are larger than the linear pre-trend difference in slopes observed in the data (+0.26 crimes per 1mi^2/week).

Section A of the Appendix additionally explores whether crime is displaced rather than reduced, finding no detectable changes for neighborhoods further than 0.25 miles from an Impact Zone. Neighborhoods within this radius report slight increases in non-major felonies and violent misdemeanors. However, when we expand treatment to include neighboring tracts within 0.25 miles of an Impact Zone, we estimate statistically significant reductions in felonies and violent misdemeanors as well as major felonies.

V Effect on High School Dropout Rates

We are interested in estimating impacts of stops on high school dropouts since high school students face critical educational decisions and are one of the most stopped demographic groups overall (see Figure A.2).[34] We proxy for exposure to stops in a student's peer network by exploiting school-level variation in stop exposure. The advantage of this approach is that it is able to detect effects in the presence of within-school peer effects and accounts for the fact that high school students likely spend time in areas outside of their home Census tract.[35]

To this end, we split schools into quartiles based on the number of stops per square mile

[32]This figure is obtained by adding Mello's IV estimate on violent crime (4.27) and property crime (15.39), since these categories collectively cover the seven offenses in our major felony category.

[33]Formally, Roth and Rambachan (2021) reports confidence sets on DD coefficients when allowing for violations of slope M between any two consecutive periods.

[34]High school students are 21 times more likely to be stopped than middle school students, and Black eighteen year-old residents in particular, face the highest stop rate of any sex-by-age demographic group.

[35]In Section B the Appendix, we report similar but more muted results using student-level variation in neighborhood stop exposure that does not account for peer effects. Unfortunately, we lack data that identifies which individual students are stopped by police.

observed in students' home Census tracts during an earlier training period. More specifically, this process involves the following steps. First, we record the average number of stops per square mile observed for a given Census tract from 2007 to 2009. Then, we link these training-period stop rates to the Census tracts of students attending a given school during the sample period, where we have valid home tracts for over 97% of students. Finally, we calculate average training-period stop rates across all students attending a given school from 2010 to 2017 and split schools into student-weighted quartiles, such that each school quartile contains approximately 25% of the student population.[36]

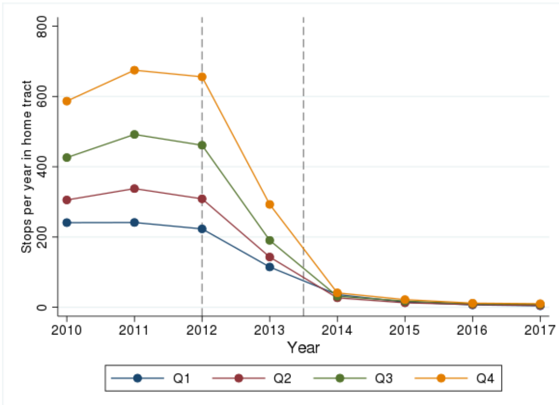
Figure VI depicts observational trends in stops and schooling outcomes by school quartile. Panel A shows that after the *Floyd* decision, the average student attending a fourth (third) quartile school is exposed to 625 (425) fewer stops per year within their home Census tract, which recall spans four-by-four blocks. In contrast, students attending below-median schools experience a reduction of about 250 stops per year in their home Census tracts, on average. About 20-25% of these stops would have involved 14 to 18 year-old residents. Panel D confirms that, within each school quartile, felony crimes per square mile remained relatively flat throughout the entire study period.

Panels E and F display raw outcome means by school quartile. Simple time-series breaks observed in Panel E suggest that the probability a student was discharged by institutional directive declined by 0.15 percentage points per year (or 60% of the pre-period mean) for students attending fourth-quartile schools, and declined by 0.10 percentage points per year (or 75%) for students attending third-quartile schools. Smaller declines are also observable for bottom-half schools, although trends in these schools are considerably noiser. Panel F displays raw changes in dropout rates. While students in third and fourth quartile schools experience sharp declines in 2014, increases in 2013 across all school quartiles complicate a simple time-series analysis. We resolve this issue by implementing a difference-in-differences strategy that compares outcome changes in fourth and third quartile schools to outcome changes in below-median "control" schools. This approach underestimates treatment effects since effects on students in control schools are differenced out.

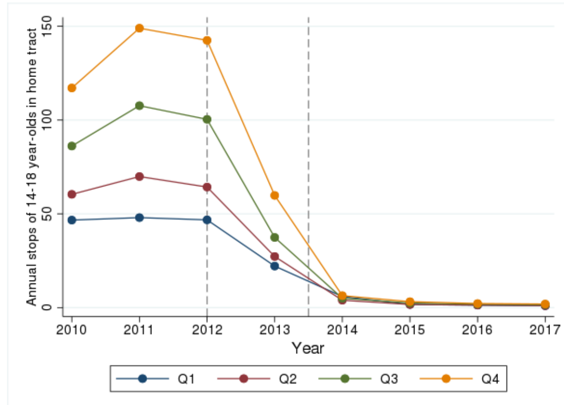
[36]Note that this approach weights training-period stop rates by the endogenous composition of student home addresses observed during the sample period. This could introduce bias if changes in stops differentially affect family location decisions across school quartiles. However, we did not find this to be the case when use tract weights from 2009 and run our main regressions on the sub-sample of schools that are operational from 2009-2017.

Figure VI: Raw Means of Stops, Felonies, and HS Outcomes by School Quartile

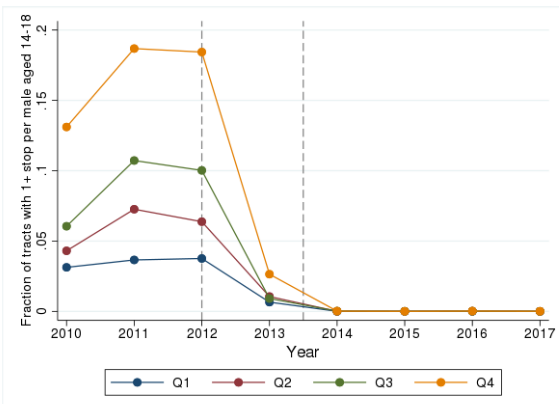
(a) Stops per Year in Students' Home Census Tract



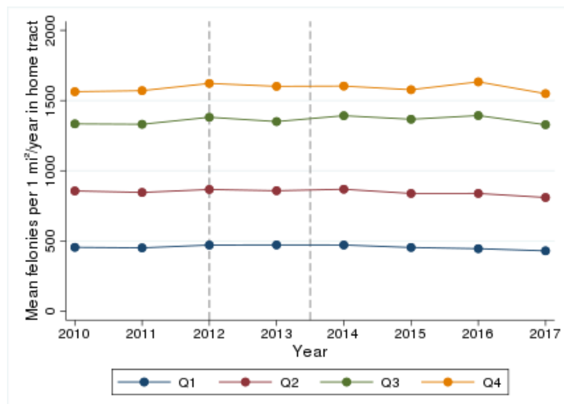
(b) Stops per year of Pedestrians Aged 14-18 in Students' Home Census Tracts



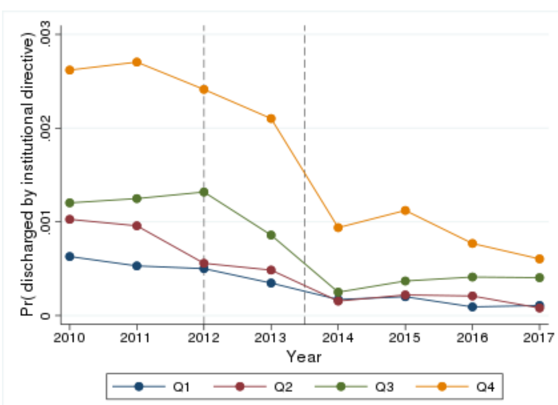
(c) Fraction of Students' Home Census Tracts with 1+ Stop per Male, Aged 14-18



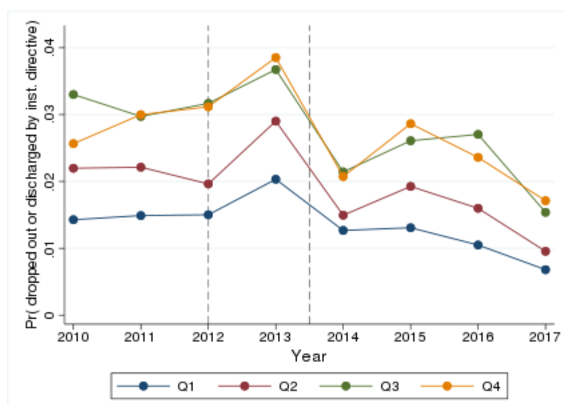
(d) Felonies per 1 mile^2/year in Students' Home Census Tract



(e) Pr(Discharged by Institutional Directive)



(f) Pr(Dropped Out or Discharged by Inst. Dir.)



Notes: Panels A and B display raw means of stops and felonies per 1 mile^2/year in the home Census tract of students attending a given school. We report means by school quartile, where quartiles are generated based on mean stops per 1^2 mile experienced in student home Census tracts during school years 2006-2007 through 2008-2009. Panel C graphs the mean likelihood of enrolled students being discharged from high school by institutional directive during expected grades 9 through 12. Panel D graphs the mean likelihood of enrolled students dropping out or being discharged by institutional directive.

A Empirical Strategy

We estimate the contemporaneous impacts of *Floyd* on dropouts and institutional discharges using a differences-in-differences design. Recall that our outcomes of interest equal one if and only if a student displays a change to their enrollment status during the course of *a given academic year*. This allows us to isolate the precise timing of when school enrollment outcomes were affected by the reform. On our student panel, we run the following estimating equation:

$$Y_{isy} = \sum_{q=3}^{4} \sum_{\tau \neq 0} \beta_{q\tau} I(q)_{\tau} + \delta_s + \delta_{g \text{ x } y} + \delta_{b \text{ x } y} + \Gamma X_i' + \epsilon_{isy} \tag{6}$$

Y_{isy} is the outcome indicator for individual i attending school s in year y. A student's expected grade level is denoted by g and is measured from the student's initial enrollment in 9th grade. δ_s are school fixed effects and $\delta_{g \text{ x } y}$ are expected-grade-by-year fixed effects, which control for changes in outcomes as students progress through high school. $\delta_{b \text{ x } y}$ are borough-by-year fixed effects that restricts comparisons to be between students attending schools in the same borough. X_i' is a vector of time-invariant student characteristics that include quintiles of age upon entering 9th grade, Black and Hispanic race indicators, a male indicator, and separate indicators for whether the student speaks Spanish or another non-English language at home. $I(q)_{\tau}$ are relative time to treatment indicators, which equal one for students attending schools in stop quartile q if year y is τ years from the onset of *Floyd*. We set the reference group as schools with below-median exposure to stops during the training period. The coefficients of interest $\{\beta_{q,t}\}$ then represent the average change between year τ and the year prior to *Floyd* among students in quartile q schools relative to the same change over time among students in below-median schools in the same borough b. All standard errors are clustered at the school-cohort level.

To ensure that differences in estimated effects are not driven by variation in pre-treatment trends, we also report results for a linear trend-adjusted version of Equation 6:

$$Y_{isy} = \sum_{q=3}^{4} \sum_{\tau=1}^{5} \beta_{q\tau} I(q)_{\tau} + \sum_{q=3}^{4} [I(q) \text{ x } \tau] + \delta_s + \delta_{g \text{ x } y} + \delta_{b \text{ x } y} + \Gamma X_i' + \epsilon_{isy} \tag{7}$$

Equation 7 is equivalent to Equation 6 except that event-time-quartile indicators are dropped for $\tau < 1$ and interactions of school quartiles with linear event-time are added for third and fourth quartile schools (i.e. $\sum_{q=3}^{4} [I(q) \text{ x } \tau]$).

While we primarily focus on reduced-form impacts of the reform, we also estimate effects per reduction in stops. To this end, we instrument for neighborhood stops per year with a treatment indicator set to one in the post-decision period (2014−2017). The reference period spans 2010 to 2012. Using a two-stage least squares regression, the estimating equation is then:

$$Y_{isy} = \sum_{q=3}^{4} \gamma_s^q Stops_{iy} + \sum_{q=3}^{4} \beta_{q1} I(q)_{\tau=1} + \delta_s + \delta_{g \text{ x } y} + \delta_{b \text{ x } y} + \Gamma X_i' + \epsilon_{isy} \tag{8}$$

where $\{\gamma_s^q\}$ estimate the effect of changes to neighborhood stops on outcomes for students attending schools in quartile q. Adding linear pre-trend controls, the equivalent regression becomes:

$$Y_{isy} = \sum_{q=3}^{4} \gamma_s^q Stops_{iy} + \sum_{q=3}^{4} [I(q)_{\tau=1} + I(q) \text{ x } \tau] + \delta_s + \delta_{g \text{ x } y} + \delta_{b \text{ x } y} + \Gamma X_i' + \epsilon_{isy} \tag{9}$$

B Results

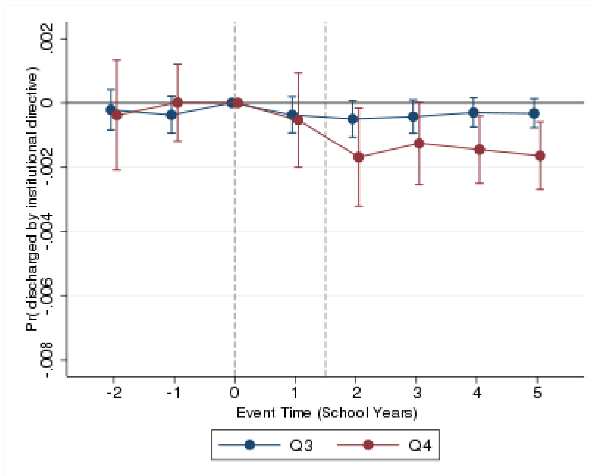
Our main findings are reported in Table VII. Following the *Floyd* decision, students in fourth (third) quartile schools experienced 377 (177) fewer stops per year, relative to students in control schools; 82 (36) of these prevented stops would have been of 14 to 18 year-old residents.

Figure VII graphs reduced-form estimates of the impact of stop reductions on educational outcomes using Equations 6 and 7. The omitted period is the 2012 school year. For our first outcome – the likelihood that a student is discharged by institutional directive in a given school year – there is little evidence of differential group trends prior to *Floyd*, especially for fourth quartile schools. If anything, there is a slight positive pre-trend, leading Equation 6 to potentially underestimate treatment effects. Following the *Floyd* decision, we estimate that students in fourth quartile schools were 0.14 percentage points or 54% less likely to leave high school during a given school year by institutional directive ($p = 0.004$). In other words, the reform prevented 84 students per year from being discharged by institutional directive in fourth quartile schools. Adjusting for linear pre-trend differences across study groups increases the estimated reduction to 0.22 percentage points (85%), but this estimate is no longer statistically significant.

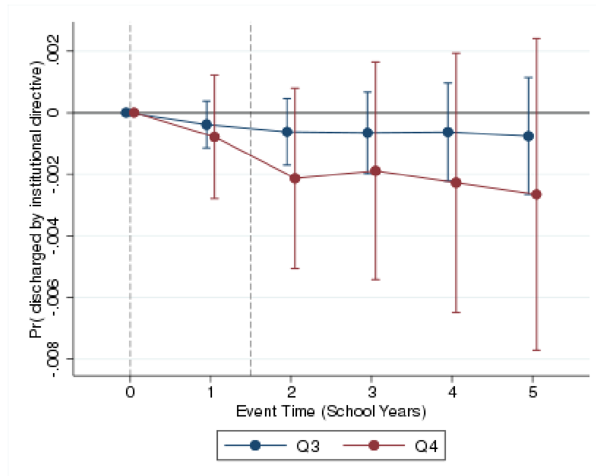
Figure VII: Effect on School Discharges

DV = Pr(Discharged by Institutional Directive)

(a) Difference-in-differences

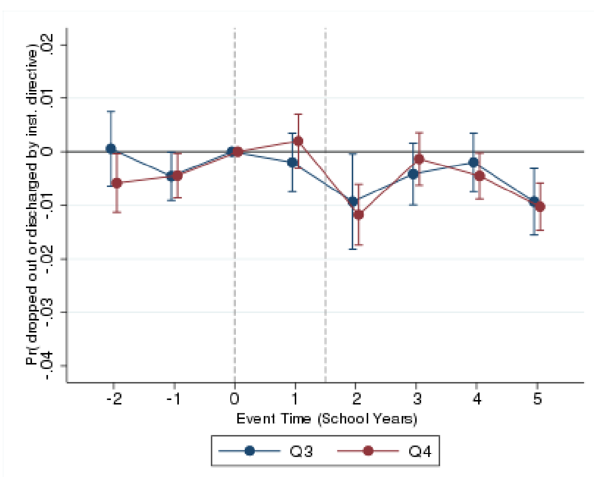


(b) + Linear Pre-trend Controls

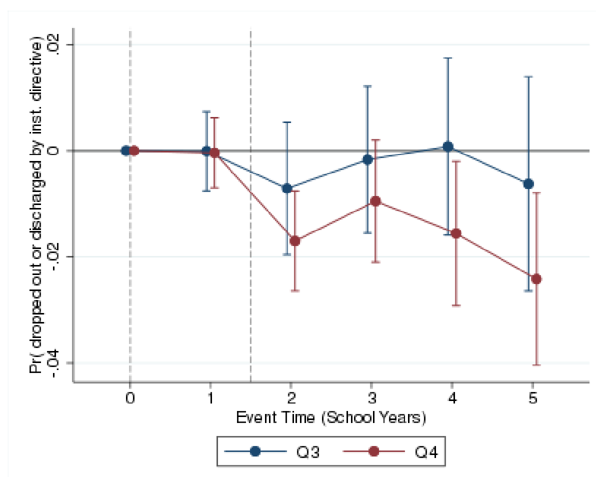


DV = Pr(Dropped Out or Discharged by Inst. Directive)

(c) Difference-in-differences



(d) + Linear Pre-trend Controls



Notes: This figure graphs coefficients from Equation 6 in Panels A and C and coefficients from Equation 7 in Panels B and D. Panels A and B display effects on the likelihood of enrolled students being discharged from high school by institutional directive during expected grades 9 through 12. Panels C and D display impacts on the likelihood of enrolled students dropping out or being discharged by institutional directive. The reference group is set to schools with below median exposure to stops during the training period. Blue dots refer to point estimates for schools in the third quartile, while maroon does denote point estimates for schools in the fourth and "most exposed" quartile. Whiskers mark 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the school-cohort level.

Our second outcome examines the holistic impact of *Floyd* on the likelihood that a student drops out or is discharged by institutional directive. Panel C of Figure VII reveals that both third and fourth-quartile schools display upward pre-trends relative to control schools. Even so, our difference-in-differences approach still finds statistically significant reductions in dropout rates following the *Floyd* decision. Students from fourth (third) quartile schools are 0.36 (0.49) percentage points or 12% (16%) less likely to drop out or be discharged by institutional directive in a given school year. Panel D of Figure VII shows that once we control for linear pre-trend differences, drop out rates are virtually unaffected during the scale-down year (2013), but then fall by about 60% for students in fourth quartile schools following the *Floyd* decision and remain low for the subsequent three school years. As shown in Table VII, the likelihood that a student dropped out or was discharged by institutional directive fell by 1.66 percentage points per school year (58%) during the post-ruling period in fourth-quartile schools relative to control schools ($p = 0.008$). Averaging point estimates across specifications, we estimate that the reform prevented 660 high school students per year from dropping out or being discharged by institutional directive. Point estimates for third-quartile schools, when controlling for linear pre-trends, are negative but considerably smaller and statistically insignificant.

To put these effects in context, Bacher-Hicks and de la Campa (2020) use a precinct commander movers design to show that students attending middle schools in precincts with an additional 350 stops per year exhibit a 0.4 percentage point increase in the likelihood of dropping out from high school. Multiplying our per-year estimate by four years of high school, the equivalent statistic for students in fourth-quartile schools is 4.0 percentage points.[37] It is not surprising that we document substantially larger treatment effects given that our approach differs in a number of key ways. First, we focus on student exposure in high school rather than during middle school. Fourteen to eighteen-year-old residents are over twenty times more likely to be stopped by police than eleven to thirteen-year-old residents. Second, we are able to leverage larger absolute differences in stop exposure. Students in fourth quartile schools experienced 377 fewer stops per home Census tract. Given there are approximately 28 tracts per precinct, our first stage involves a change in neighborhood exposure to stops that is about 30 times larger (in terms of stops per area). Finally, our approach is better able to capture neighborhood stop exposure within a student's peer network. While (Bacher-Hicks and de la Campa, 2020) measured stop exposure by summing stops over school district geographies, we find that our main effects persist even when limiting treatment variation to occur between schools *within the same school district*.[38] This highlights the importance of

[37]This assumes an effect size of 1.01 percentage points, which is the average across our two specifications.

[38]Inclusion of school district fixed effects does little to affect our point estimates. These results are

Table VII: Effect of *Floyd* on High School Dropout and Institutional Discharge Rates

	Pre-period	Base		+ Linear Pre-trends	
	Treat Mean	$\beta_{\tau\geq 2}$	P-value	$E[\beta_{\tau\in[2,5]}]$	P-value
	(1)	(2)	(3)	(4)	(5)
DV = Stops per year					
Third Quartile	459.863	-176.720***	0.000	-188.163***	0.000
		(5.191)			
Fourth Quartile	639.028	-376.543***	0.000	-475.060***	0.000
		(5.012)			
DV = Stops of 14-18 year-old residents per year					
Third Quartile	98.094	-35.675***	0.000	-47.812***	0.000
		(1.342)			
Fourth Quartile	136.129	-81.766***	0.000	-123.255***	0.000
		(1.228)			
DV = Pr(Discharged by Institutional Directive)					
Third Quartile	0.0013	-0.0002	0.261	-0.0007	0.368
		(0.0002)			
Fourth Quartile	0.0026	-0.0014***	0.004	-0.0022	0.260
		(0.0005)			
DV = Pr(Dropped Out or Discharged by Inst. Directive)					
Third Quartile	0.0315	-0.0049**	0.029	-0.0036	0.638
		(0.0022)			
Fourth Quartile	0.0289	-0.0036**	0.020	-0.0166***	0.008
		(0.0015)			

Notes: This table reports coefficients from Equations 6 and 7. Treatment is defined as schools with students from neighborhoods with stop exposure during the training period that ranked between the 50th and 75th percentile – "third quartile" – or 75-100th percentile – fourth quartile. Column 1 reports mean outcomes for each quartile during the 2010 through 2012 school years. Column 2 reports the coefficient on a post-ruling indicator that collapses 2014-2017 treatment indicators in Equation 6 into a singular indicator. Standard errors are reported below coefficients in parentheses. Column 4 reports the average across all four post-ruling coefficients from Equation 7, and Column 5 reports the p-value from an F-test that the four post-ruling coefficients are jointly zero. Standard errors are clustered at the school-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

using detailed information on where students and their peers live to capture neighborhood stop exposure.[39]

How do our treatment effects compare to those observed for students who are exposed to a deadly police shooting? Ang (2021) finds that high school students living within a 0.5-mile radius of a police shooting were 1.1 percentage points less likely to graduate from high school relative to students living in the same Census block group but further from the shooting.[40] Comparatively, treatment effects observed among students attending fourth-quartile schools are almost four times larger. In terms of students affected, they are about 10 times larger.

Table VIII: IV Estimates of Stops on High School Dropout and Institutional Discharge Rates

	Pre-period	Base		+ Linear Pre-trends	
	Treat Mean	100 Stops	P-value	100 Stops	P-value
	(1)	(2)	(3)	(4)	(5)
DV = Pr (Discharged by Institutional Directive) x 100					
Third Quartile	0.126	0.011	0.260	0.032	0.164
		(0.010)		(0.023)	
Fourth Quartile	0.258	0.037***	0.004	0.042	0.135
		(0.013)		(0.028)	
DV = Pr (Dropped Out or Discharged by Inst. Directive) x 100					
Third Quartile	3.147	0.275**	0.027	0.282	0.405
		(0.124)		(0.339)	
Fourth Quartile	2.893	0.095***	0.019	0.219***	0.026
		(0.041)		(0.098)	

Notes: This table estimates the impact of 100 stops per year in a student's home Census tract on high school dropout and institutional discharge rates. The outcome is in percentage points to ease interpretation. Coefficient are estimated using Equation 8 and 9. Treatment is defined as schools with students from neighborhoods with stop exposure during the training period that ranked between the 50th and 75th percentile – "third quartile" – or 75-100th percentile – fourth quartile. Column 1 reports mean outcomes for each quartile during the 2010 through 2012 school years. Column 2 instruments for post-ruling changes in stops with a singular indicator that equals one for treatment groups in school years 2014-2017. Column 4 instruments for post-ruling changes in stops with year indicators for years 2014-2017, controlling for pre-trend differences in trends. Standard errors are clustered at the school-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table VIII, we report coefficients from our IV regressions. In fourth quartile schools, reducing average stop rates across students' home neighborhoods by 100 stops leads to a

available upon request.

[39]Future drafts will broaden our measure of stop exposure to include exposure to stops around one's high school.

[40]Students who observed a police killing in 9th or 10th grade were the most affected – exhibiting a 1.8 percentage point decline in graduation rates.

0.037 percentage point reduction in the probability a student is institutionally discharged in a given school year ($p = 0.004$). For these same students, the probability of dropping out or being institutionally discharged falls by 0.095 to 0.22 percentage points per year ($p < 0.03$), and is 0.28 percentage points for students in third quartile schools ($p = 0.027$). Given that the net present value of graduating high school is estimated to be around $300,000, these effect sizes are economically large (Vining and Weimer, 2019). That is, our IV estimates suggest that conducting an average of 100 additional stops per year in the two-block radius of where students live carries a social cost of about $23 million due to spillovers on high school graduation rates alone.[41] These findings highlight just how harmful frequent stop encounters can be to the life trajectory of high school students.

Robustness

It is possible that the timing of *Floyd* coincided with other educational reforms that affect our outcomes of interest. Most prominent of which were changes to school discipline policy. During the 2013 school year, the Department of Education removed low-level suspensions for disorderly behavior, such as profane language or persistent non-compliance. Instead of being suspended for 1 to 5 days, less disruptive punishments were adopted, such as removing students from a single class. Craig and Martin (2019) study this reform by comparing middle schools that prior to the reform displayed high versus low suspension rates. They find the reform increased math (reading) scores of middle school students by 0.05 (0.03) standard deviations. To test whether our findings can be explained by this suspension reform, we split schools into high (above-median) and low (below-median) suspension high schools using mean disorderly suspension rates observed from 2008 and 2009. We adapt Equations 6 and 7 by fulling interacting an "above-median" stop-exposure indicator with an above-median suspension indicator. This approach compares outcome changes in high-stop high-suspension schools, high-stop low-suspension schools, and low-stop high-suspension schools to low-stop low-suspension schools. Figure A.6 displays difference-in-differences coefficients and Table A.6 reports post-ruling coefficients.

Figure A.6 shows that reductions in suspension rates sharply change in 2013, while changes in dropouts and institutional discharges bottom out in 2014. In terms of levels, Panel B shows that high-stop high-suspension schools display the largest declines in institutional discharges, suggesting that reductions in disorderly suspensions may have complemented reduced stop rates. However, high-stop low-suspension schools also display substantial and statistically significant declines, especially when viewed as a percent of pre-reform mean; the post-ruling difference-in-differences coefficient implies a 32% decline ($p = 0.007$). What's

[41] This estimate assumes 18% of students who dropout or are discharged by institutional directive will go on to graduate in six years, and assumes a conservative effect size of 0.1575 percentage points.

more, dropout rates in high-stop low-suspension high schools closely track those of high-stop high-suspension schools, while low-stop high-suspension schools do not. Therefore, we argue that the simultaneous change to disorderly suspension policy may have complemented the reform but is unlikely to be the driving force behind our main findings.

A second set of disciplinary reforms was ushered in by the newly-elected Mayor Bill de Blasio and his School Chancellor Carmen Fariña. While Chancellor Fariña openly promoted less punitive discipline procedures after taking office in January 2014, it was not until February 2015 that she officially revised the school discipline code.[42] This was a full year after dropout and institutional discharge rates had bottomed out, suggesting that these reforms do not explain our findings. More broadly, there does not appear to be sharp changes to other important school inputs. For example, Appendix Figure A.11 shows that summonses and arrests made by school safety agents steadily declined from 2011 to 2016.

We also probe robustness by exploring parallel trend violations per Roth and Rambachan (2021). Figures A.7 and A.8 show that our findings are robust to fairly large violations so long as we assume pre-trend violations introduce positive rather than negative bias. Finally, Section B of the Appendix reveals that our main findings persist if we instead exploit tract-level variation in stop exposure rather than school-level variation. Students that experience larger reductions to stop rates in their home neighborhoods due to the reform also display significantly larger reductions in dropout rates, with effect sizes that are similar in magnitude to our main findings.

Heterogeneity

How do effects differ by student race and sex? It is first important to note that 92% of students attending schools in the fourth quartile are Black or Hispanic, meaning that the *Floyd* decision disproportionately benefited students of color. We formally explore whether demographic sub-populations differentially responded to the reform by running Equations 6 and 7 separately on each race, sex, and race-by-sex subgroup. To improve precision, we pool white, Asian, and other-race students into one race category.

Results on the likelihood a student is discharged by institutional directive are reported in Table A.8 and results for the likelihood a student drops out or is discharged by institutional directive are reported in Table A.9. We only report coefficients on fourth-quartile school indicators, so as to reduce the number of hypotheses tested. Point estimates across both specifications suggest that impacts on discharges by institutional directive are slightly larger

[42]These reforms were moderate, however, changing only the most contentious aspects of the school discipline code. Most notably, under the new policy principals needed approval from the Education Department's central office before suspending a student and the NYPD were forced to track and report any instance in which a student is handcuffed (Harris, 2015).

for black students, although differences between estimated coefficients are not statistically significant. If we average post-ruling coefficients across the two specifications, we find that the yearly likelihood of dropping out or being discharged by institutional directive falls by 1.1 percentage points (38%) per year for Black students, 0.7 percentage points (25%) per year for Hispanic students, and 0.9 percentage points (38%) per year for white/Asian/other-race students. As one might expect, these tables also reveal that male students benefit more from the *Floyd* decision than female students, especially with regard to institutional discharge rates. Black male students in particular display the largest absolute reductions. Scaling estimates by the number of students, we estimate that the reform prevented about 50 Black male students per year from becoming involved with the carceral state.

VI Welfare Implications of the *Floyd* Reform

This paper provides causal estimates of the impact of the *Floyd* ruling on crime rates and high school dropout rates. To understand the welfare impact of the reform, we follow Hendren and Sprung-Keyser (2020) in estimating the *marginal value of public funds* (MVPF). As Finkelstein and Hendren (2020) describe, "the MVPF is the ratio of the marginal benefit of the policy to the net marginal cost to the government of the policy; crucially, this net marginal cost is inclusive of the impact of any behavioral responses to the policy on the government budget." More generally, this ratio captures benefits in dollars to the beneficiaries of the policy for every $1 of net government spending, and allows for easy comparison across policies (Hendren and Sprung-Keyser, 2020).

In the case of the *Floyd* reform, the costs to the government budget include direct costs related to policing changes and indirect costs related to criminal and educational responses to the policy change. Educational impacts on younger children and labor market impacts on adults are out of the scope of our analysis but likely result in the underestimation of the potential cost savings of this reform.

While the reform dramatically reduced stop rates by altering the rules of when a pedestrian stop is justified, it did *not* alter the number of officers deployed, or even where they were deployed. While we do not have access to linked payroll data, there is no evidence of changes in officer salaries, overtime pay, or the size of the force during our study period, especially in the years directly proceeding and following the reform (see Figure A.12). It is possible that officers shifted efforts away from conducting pedestrian stops for low-level offenses and switched their focus toward more serious offenses. Nevertheless, the outcome of such efforts should be captured in our estimates of changes to reported crime. The direct policing cost is thus taken to be zero.

Section IV reports null effects for shootings, killings, and felonies and violent misdemeanors. There is a small and marginally significant reduction in non-violent misdemeanors and violations that can be attributed to a reduction in detection of these events through stop-related arrests and ticketing. As such, a cost-weighted crime measure a la Mello (2019) yields a noisy but negative point estimate. We, therefore, make the conservative assumption of null crime responses.[43]

To estimate the social cost of educational responses, we first calculate the number of high school students who are prevented from dropping out of high school each school year. Taking the average point estimate across our two empirical strategies and scaling by the number of students in fourth-quartile schools, we estimate that the reform prevented about 660 students per year from dropping out or being discharged by institutional directive. Multiplying by the fraction of these students who will not go on to graduate high school in six years (82%) and the fiscal externality of an additional high school graduation, we estimate that the annual cost savings to the government budget among fourth-quartile students alone is $65 million (Vining and Weimer, 2019).[44]

Combining the above costs yeilds a cost savings to the government of $65 million per year, meaning that the reform pays for itself. Provided that residents' willingness to pay (WTP) is positive, then the MVPF is infinite. Neighborhood resident WTP consists of two core ingredients, their WTP for a reduction in pedestrian stops, and the WTP for higher lifetime earnings among 660 students who are prevented from dropping out of high school each year. Invoking the envelope theorem and assuming income gains stem to returns to human capital allows us to estimate the latter using the policy's impact on net income after taxes (Hendren and Sprung-Keyser, 2020), which is worth $143 million per year (Vining and Weimer, 2019). For the former, we estimate that the direct benefit to pedestrians of avoided stops constitutes an additional $3 million per year.[45] Combining, resident WTP for the reform is $146 million per year. This is a conservative estimate since it does not account for other potential benefits to residents, such as improved educational achievement among younger students, and reduced criminal justice involvement and better labor market prospects among adult residents.

[43]We use reported crimes to measure criminal responses since these data encapsulate all crime changes, instead of changes in crime detected directly through pedestrian stops.

[44]We assume a 3% real discount rate and calculate the fiscal externality as the difference between private educational returns and total social value estimates described in Table 3 of Vining and Weimer (2019). All calculations are inflation-adjusted to be in 2022 dollars.

[45]To arrive at this figure, we set the price a pedestrian is willing to pay to avoid a stop at twice the minimum wage rate and assume the average stop takes 15 minutes. Steady-state differences in stop rates are calculated by subtracting stop rates observed three to five years after the reform from stop rates observed in the four years prior to the reform.

VII Conclusion

Through the lens of the *Floyd* reform, this paper examines the effectiveness of using police stops as a crime deterrence tool. We find that sizeable reductions in pedestrian stops have no detectable impact on serious crimes, and *decrease* the detection of minor offenses. These results contrast the effectiveness of police surges that simultaneously increased officer presence and stop rates in crime hot-spots. Together, our results suggest that *patrol officer presence rather than concentrating pedestrian stops in higher-crime areas* matter for deterring major felonies.

The concentration of stops in higher-crime neighborhoods leads to large racial disparities in stop rates, where young men of color are stopped at alarming rates. These excessive stop rates spillover into the classrooms of neighborhood high school students. We estimate that the reform prevented about 660 students per year from dropping out or being discharged by institutional directive, leading the reform to pay for itself and provide a social value of at least $210 million per year. Over ninety percent of these teenagers were Black or Hispanic. Given that our approach differences out effects for students attending less-exposed schools and that we do not account for educational effects on younger children or labor market impacts on young adults, the net social benefit of the reform is likely much larger.

This paper highlights the need for careful analysis when examining the social impact of police practices. Much of the social value of policing is not directly observed. Instead, accurate estimation of social value requires causal estimation of first-order behavioral responses, such as crime, educational, and labor market responses. Our findings illustrate the outsize role the police play in urban communities of color, and especially, for young men of color. To better understand the extent to which the criminal justice system contributes to racial disparaties, future research could examine the labor market consequences of police encounters and arrests for low-level offenses that have the potential to disrupt the life trajectory of young people. Setting policing goals, and more broadly, criminal justice policy that takes these spillover effects into account is essential for reducing existing racial and socioeconomic opportunity gaps.

References

ACLU of Illinois (2015, August). *Investigative stop and protective pat down settlement agreement.* Chicago Police Department, ACLU of Illinois, City of Chicago.

ACLU of Pennsylvania (2011, July). *Bailey, et al. v. The City of Philadelphia, et al.* U.S. District Court for the Eastern District of Pennsylvania.

ACLU of Pennsylvania (2021, June). Federal judge orders new approach to eliminate racial bias in stop and frisk policing in philadelphia.

Agan, A. Y., J. L. Doleac, and A. Harvey (2021a). Misdemeanor prosecution. *No. w28600. National Bureau of Economic Research*.

Agan, A. Y., J. L. Doleac, and A. Harvey (2021b). Misdemeanor prosecution. Technical report, National Bureau of Economic Research.

Ang, D. (2021). The effects of police violence on inner-city students. *Quarterly Journal of Economics 136*(1), 115–168.

Bacher-Hicks, A. and E. de la Campa (2020). Social costs of proactive policing: The impact of nyc's stop and frisk program on educational attainment. *Working paper.*.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Boyd, R. W. (2018). Police violence and the built harm of structural racism. *Lancet 392*, 258–259.

Braga, A. A. and B. J. Bond. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology 46*(3), 577–607.

Buchanan, L., Q. Bui, and J. Patel (2020, July). Black lives matter may be the largest movement in U.S. history. *The New York Times*.

Caetano, G. S. and V. Maheshri (2014). Identifying dynamic spillovers in criminal behavior. *Available at SSRN 2460952*.

Chalfin, A. and J. McCrary (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature 55*(1), 5–48.

Chalfin, A. and J. McCrary (2018). Are US cities underpoliced? Theory and evidence. *Review of Economics and Statistics 100*(1), 167–186.

Chetty, Raj, H. N. and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*, 855–902.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020). Race and economic opportunity in the united states: An intergenerational perspective. *Quarterly Journal of Economics 135*, 711–783.

Craig, A. C. and D. C. Martin (2019). Discipline reform, school culture, and student achievement. *Unpublished working paper*.

Dept. of Justice (2016, March). *The United States of America v. The City of Ferguson.* U.S. District Court for the Eastern District of Missouri.

Di Tella, R. and E. Schargrodsky (2004). "do police reduce crime? Estimates using the allocation of police forces after a terrorist attack.". *American Economic Review 94*(1), 115–133.

Eterno, J. A. and E. B. Silverman (2019). *The crime numbers game: Management by manipulation.* CRC Press.

Evans, W. N. and E. G. Owens (2007). COPS and Crime. *Journal of Public Economics 91*(1-2), 181–201.

Fan, J., I. Gijbels, T.-C. Hu, and L.-S. Huang (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, 113–127.

Finkelstein, A. and N. Hendren (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives 34*(4), 146–67.

Geller, A., J. F. Fagan, T. Tyler, and B. Link (2014). Aggressive policing and the mental health of young urban men. *American Journal of Public Health 104*(12), 2321–2327.

Golden, M. and C. Almo (2004). Reducing gun violence: An overview of New York City's strategies. *Vera Institute of Justice*.

Goncalves, F. and S. Mello (2021). A few bad apples? Racial bias in policing. *American Economic Review 111*(5), 1406–41.

Harrell, E. and E. Davis (2020). Contacts between police and the public, 2018 – statistical tables. Bureau of Justice Statistics, U.S. Department of Justice.

Harris, E. A. (2015, February). Suspension rules altered in New York City's revision of school discipline code. *The New York Times*.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Kirk, D. S. and A. V. Papachristos (2011). Cultural mechanisms and the persistence of neighborhood violence. *American journal of sociology 116*(4), 1190–1233.

Knox, Dean, W. L. and J. Mummolo (2020). Administrative records mask racially biased policing. *American Political Science Review 114*(3), 619–637.

Legewie, J. and J. Fagan (2019). Aggressive policing and the educational performance of minority youth. *Sociological Review 84*(1), 220–247.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives 18*(1), 163–190.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review 103*(3), 226–31.

MacDonald, J., J. Fagan, and A. Geller (2016). The effects of local police surges on crime and arrests in new york city. *PLoS one 11*(6), e0157223.

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics 172*, 174–200.

Mello, S. (2021). Fines and financial wellbeing. *Working paper.*.

National Academies of Sciences Engineering and Medicine (2018). *Proactive Policing: Effects on Crime and Communities*. Washington, DC: The National Academies Press.

Palmer, C. J., P. A. Pathak, et al. (2017). Gentrification and the amenity value of crime reductions: Evidence from rent deregulation. Technical report, National Bureau of Economic Research.

Roth, J. and A. Rambachan (2021). An honest approach to parallel trends. *Working paper*.

Vining, A. R. and D. L. Weimer (2019). The value of high school graduation in the united states: Per-person shadow price estimates for use in cost–benefit analysis. *Administrative Sciences 9*(4), 81.

Weisburd, D., L. A. Wyckoff, J. Ready, J. E. Eck, J. C. Hinkle, and F. Gajewski (2006). Does crime just move around the corner? a controlled study of spatial displacement and diffusion of crime control benefits. *Criminology 44*(3), 549–92.

Zimroth, P. (2016). Submission of second report of the independent monitor. *Southern District of New York*.

Appendices

A Robustness of Crime Responses to Police Surges

We probe the robustness of estimated crime effects of police surges in a number of ways. First, it is possible that Operation Impact displaces crime rather than deters it. To get a sense of geographic spillovers to nearby Census tracts, we explore the impact of becoming a "neighbor" of an Impact Zone on neighborhood crime rates in Appendix Table A.4. Columns 1 to 5 define treatment as neighborhoods that have $\leq$5% of their area covered by an Impact Zone and are within 0.25-miles of an Impact Zone. Event time is then defined in six-month periods relative to the time when neighborhood n switches from being outside a 0.25-mile radius of an Impact Zone to being within 0.25-mile radius. We include indicators that control for whether a neighbor is part of an Impact Zone during other assignment periods and whether non-neighbors are part of an Impact Zone. Coefficients on measures of neighborhood policing confirm that policing intensity is unaffected by nearby Impact Zone assignment. We find no evidence of spillover effects for major felonies, shootings, killings or non-violent misdemeanors and violations. However, there are small increases in non-major felonies and violent misdemeanors that are jointly significant at the 10% level. This may be the result of crime spillovers or be due to increased detection or crime reporting. In Columns 6 to 10, we repeat the same analysis except that treatment is defined as Census tracts that are within a 0.25-1.0 mile radius of an active Impact Zone (and were previously more than 1.0 mile away). Point estimates are close to zero and statistically insignificant, suggesting that crime does not spillover to neighborhoods situated more than 0.25 miles from an Impact Zone.

Given some evidence of spillovers to neighborhoods within a 0.25-mile radius of Impact Zones, Appendix Table A.5 reports results when redefining treatment as an indicator for whether a neighborhood is within a 0.25-mile radius of an active Impact Zone (including those that are partially covered by an Impact Zone). Compared to our main specification, point estimates are similar in direction but are smaller in magnitude. Post-assignment coefficients on major felonies are still around -10% of the pre-period mean and are jointly significant ($p = 0.017$). The same is true for non-major felonies and violent misdemeanors, which increase by about 4% of the pre-period mean ($p = 0.030$). Non-violent misdemeanors and violations increase initially by 3% only to taper off in years two and three. All other crime coefficients are small and jointly insignificant during the post-period.

Second, we redefine treatment as the first time a Census tract becomes part of Impact Zone (i.e. has $> 5\%$ of its area covered) during our study period. While this reduces the number of assignment events, it removes selection into treatment during the post-period, since event-time is no longer dependent on police management continuing to designate the tract as an Impact Zone after the initial six-month assignment period. Results are reported in Appendix Table A.3. Across most crime outcomes, point estimates are similar to our main results in terms of direction and magnitude, although standard errors are considerably larger. We interpret the similarity of these results as validation of our primary specification.

Finally, Appendix Figure A.5 tests robustness of effects on felonies and violent misdemeanors to parallel trends violations a la Roth and Rambachan (2021). Consistent with our trend-adjusted specification, we find that standard difference-in-difference estimates (given in blue) underestimate crime deterrence effects of Impact Zone assignment due to a positive treatment-control differences in pre-trend slopes. Once we allow for parallel trend violations, confidence sets around first-year estimates are centered around zero. Confidence intervals of second and third-year coefficients fall well below zero. Panels B and C reveal that the "break-down" slope for year two and year three estimates, respectively, is 0.35 and 0.45 crimes per 1mi^2/week. During the pre-period, Impact Zones display a differential slope of +0.26 crimes per 1mi^2/week, substantiating the claim that these police surges significantly reduced serious crime rates.

B Estimating Effects on High School Dropout Rates using Student-level Variation

As an alternative approach for estimating the impact of the reform on high school dropout rates, we compare students living in neighborhoods that prior to the reform had substantially different stop rates. To this end, we assign each student-year observation in our main sample to a stop-exposure quartile based on the mean stop rate observed for the home Census tract of a given student during the training period (i.e., from 2007 to 2009). We then estimate the following difference-in-differences (DD) regression:

$$Y_{iny} = \sum_{q=3}^{4} \sum_{\tau \neq 0} \beta_{q\tau} I(q)_\tau + \delta_n + \delta_{g \text{ x } y} + \delta_{b \text{ x } y} + \Gamma X_i' + \epsilon_{iny} \tag{10}$$

Y_{iny} is the outcome indicator for individual i living in neighborhood n in year y. Note that this specification is the same as Equation 6 except that (i) quartile q is now defined at the tract-level (and not the school-level) based on where students live in a given school year, and (ii) school fixed effects are replaced with tract fixed effects. We also cluster standard errors at the tract-by-cohort level instead of at the school-by-cohort level. If students frequently encounter police outside of their home neighborhood and/or the school concentration of stop exposure is a particularly important mediator of observed effects (e.g., via peer effects), then this approach will underestimate reductions in dropout rates.

Results are reported in Table A.7. Relative to students living in low-exposure neighborhoods, students living in fourth (third) quartile neighborhoods experience 763 (258) fewer stops per year following the *Floyd* decision. These effect sizes are almost twice as large as our main "first-stage" estimates. However, it is important to note that tract-level differences likely overstate true differences in exposure, since students do not spend all of their time in their home Census tract. Using Equation 10, we find evidence of substantial declines in both outcomes for students living in both third- and fourth-quartile neighborhoods, all of which are significant at the 1% level. Effects sizes are similar in magnitude to our main estimates, but are about half as large when controlling for linear pre-trend differences. DD coefficients are displayed in Figure A.10 and closely resemble DD coefficients from our main analysis. While we prefer the school-level analysis since it better captures stop exposure in a student's peer network, these findings confirm that our main findings persist when restricting the analysis to comparisons between students that experienced differently-sized shocks to stop rates in their home neighborhoods.

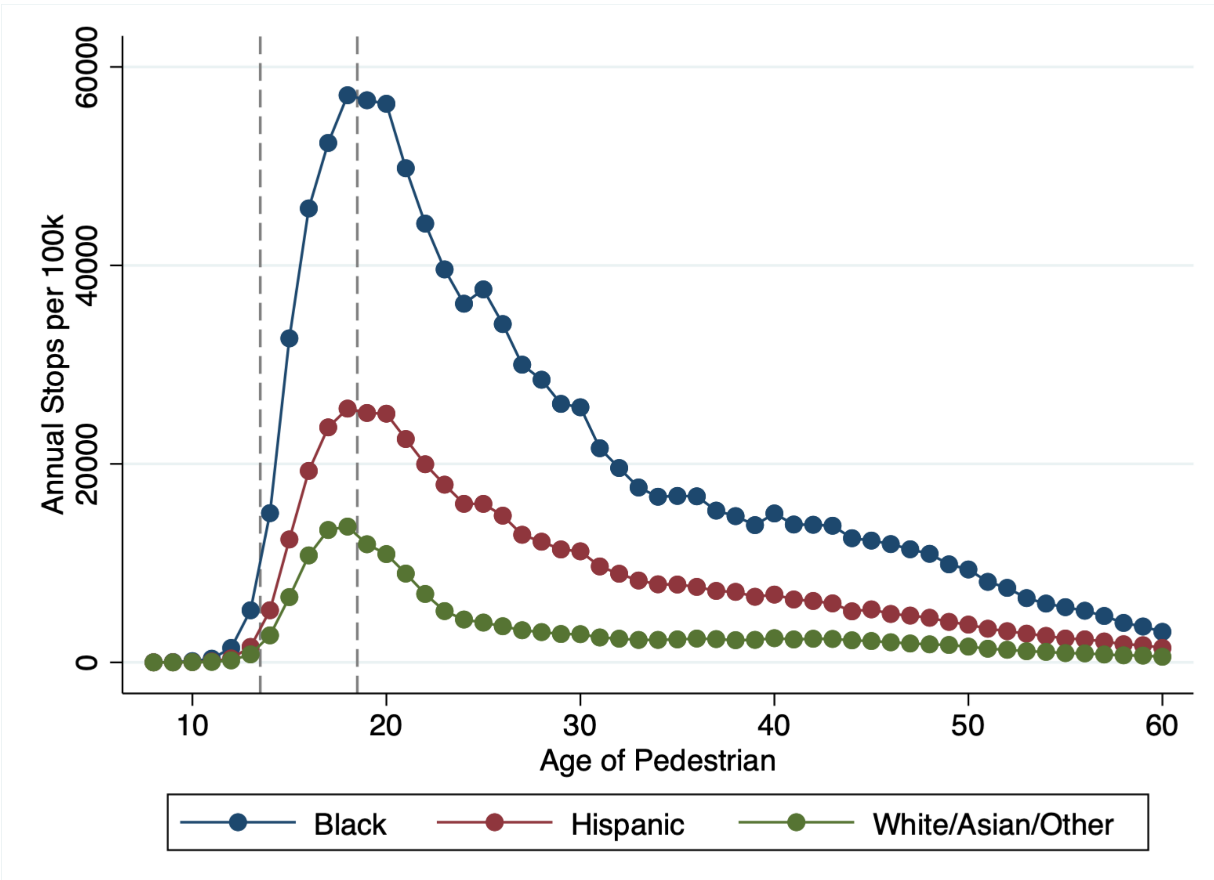
C Supplemental Figures

Figure A.1: Timeline of NYPD Stop and Frisk Policies

Jan 1, 2003	The NYPD launches Operation Impact, deploying a majority of its Academy's graduating officers to twenty Impact Zones throughout the city.
Jan 31, 2008	*Floyd*: The Center for Constitutional Rights (CCR) files initial complaint.
Sept 10, 2008	*Floyd*: Court orders the City and NYPD to turn over to CCR all stop-and-frisk data for past decade.
Oct 17, 2011	Commissioner Kelly issues department-wide order that formalizes use of stop quotas in officer performance reviews.[1]
Nov 23, 2011	Judge Scheindlin grants motion to reconsider *Floyd* case in federal district court.
Apr 16, 2012	*Floyd*: Court mostly denies City's motion to exclude Dr. Fagan's expert testimony.
May 16, 2012	*Floyd*: Court allows "Class Action" status. Commissioner Kelly issues memo scaling back stops.
Jan 8, 2012	*Ligon*: Court grants plaintiffs' motion for a preliminary injunction ordering the NYPD to immediately cease its practice of unlawful trespass stops outside Clean Halls buildings in the Bronx.
Aug 12, 2013	*Floyd*: Judge Scheindlin finds the NYPD liable for a pattern and practice of racial profiling and unconstitutional stops, and orders broad reforms to be supervised by a court-appointed monitor.
Jan 30, 2014	*Floyd*: Newly-elected Mayor de Blasio accepts Court's ruling and withdraws appeals.

Notes: This figure displays a timeline of important changes to stop and frisk policy. *Floyd* and *Ligon* refer to *Floyd et al. v. The City of New York* and *Ligon et al. v. The City of New York*, respectively. Both lawsuits were filed in the United States District Court for the Southern District of New York and were presided over by Judge Shira A. Scheindlin.

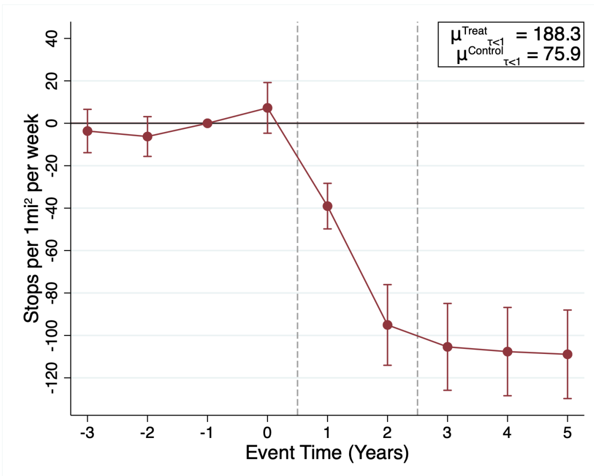
Figure A.2: Pre-*Floyd* Stop Rates by Age and Race



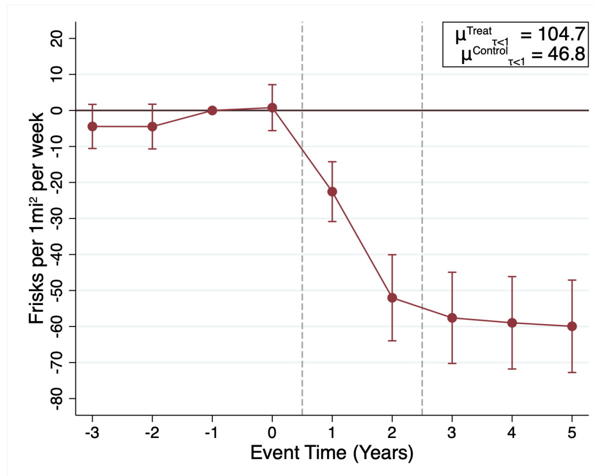
Notes: This figure displays stop rates from 2006-2011 by pedestrian age-by-race sub-groups. We use population estimates for New York City from Table SF1 of the 2010 Census, which includes population estimates for each age (under age 1 to 100) by race sub-group (e.g. All, Black Non-Hispanic, and Hispanic).

Figure A.3: *Floyd* Difference-in-differences Estimates in Highest Crime Neighborhoods

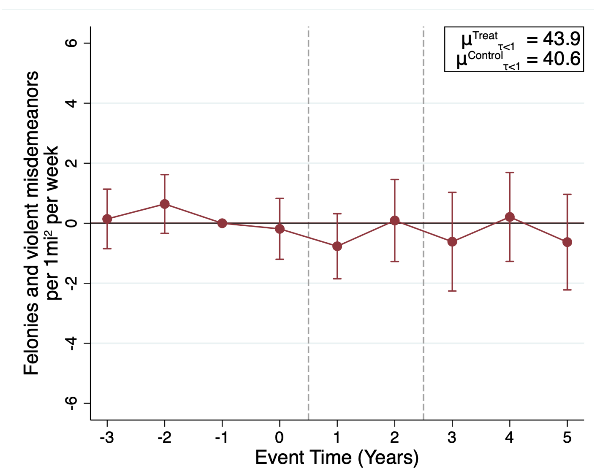
(a) Stops



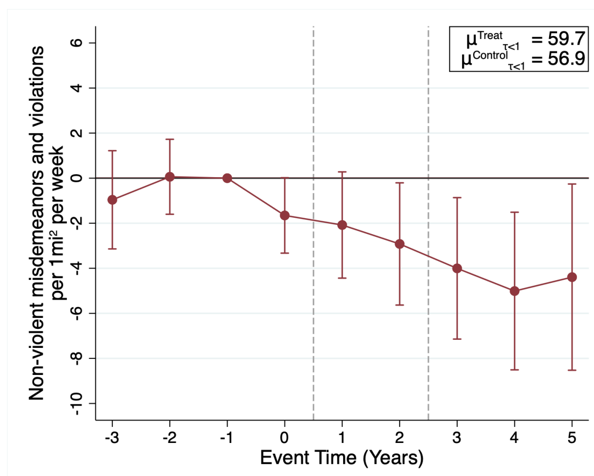
(b) Frisks



(c) Felonies & Violent Misdemeanors



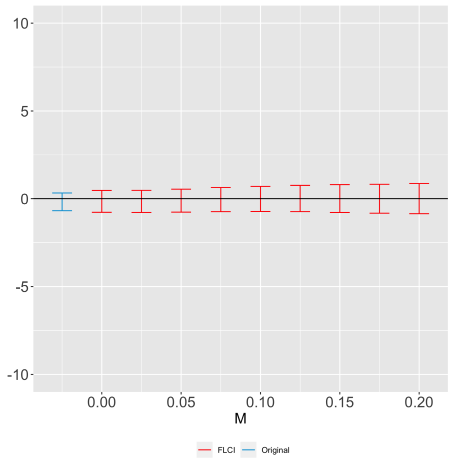
(d) Non-violent Misd. & Violations



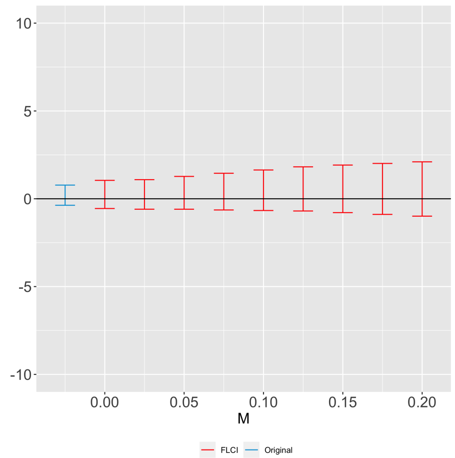
Notes: This figure graphs coefficients from Equation 3 on stops, frisks, felonies and violent misdemeanors, and non-violent misdemeanors and violations for neighborhoods that have above-median felony crime rates during the training period. Maroon dots denote point estimates and whiskers show 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the precinct level. Pre-period outcome means are given for treatment and control neighborhoods in the northeast corner of each figure.

Figure A.4: Sensitivity Analysis for DD Estimates of *Floyd* on Felonies and Violent Misdemeanors

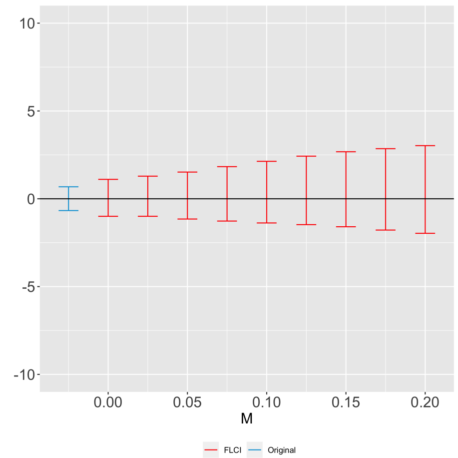
(a) $\tau = 1$



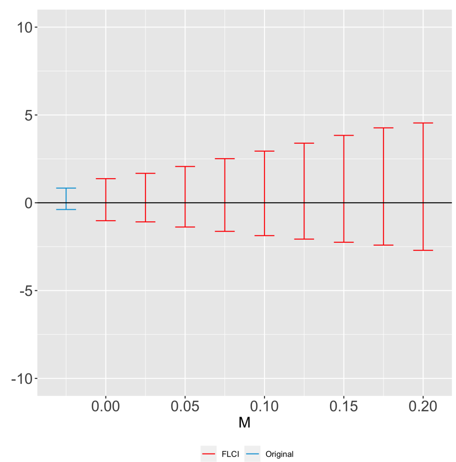
(b) $\tau = 2$



(c) $\tau = 3$



(d) $\tau = 4$



(e) $\tau = 5$

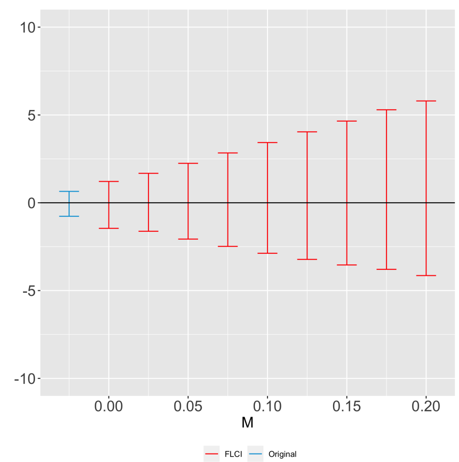
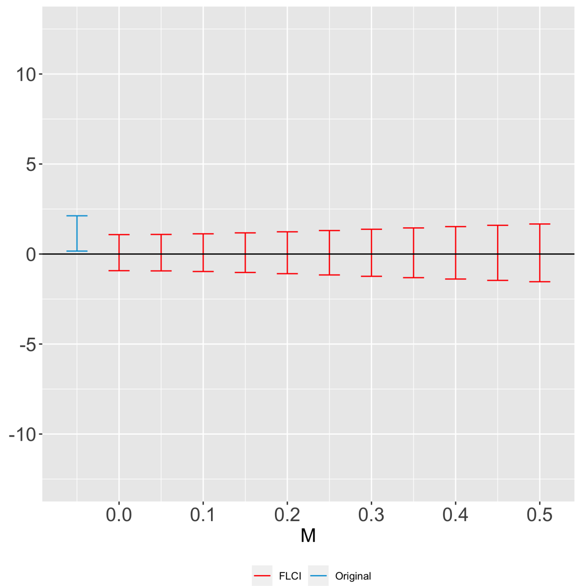
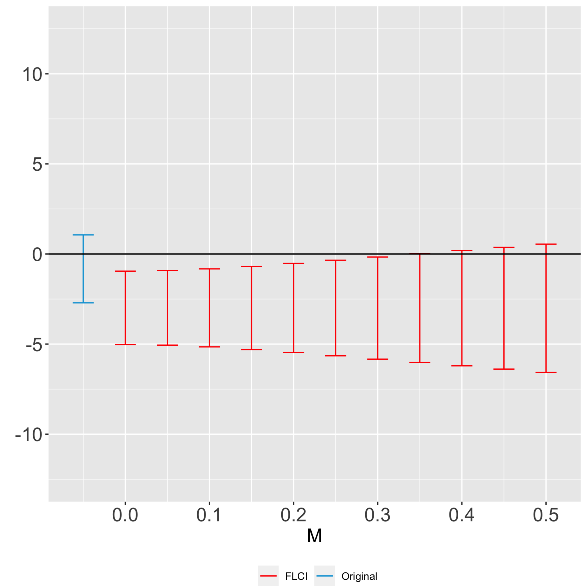


Figure A.5: Sensitivity Analysis for DD Estimates of *Operation Impact* on Felonies and Violent Misdemeanors

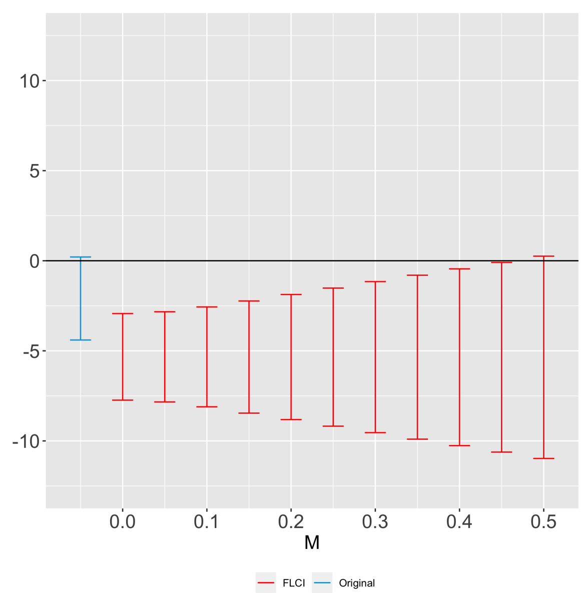
(a) $\tau = 1$



(b) $\tau = 2$



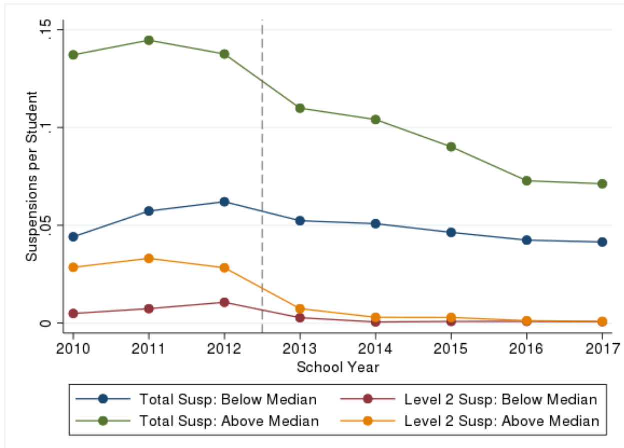
(c) $\tau = 3$



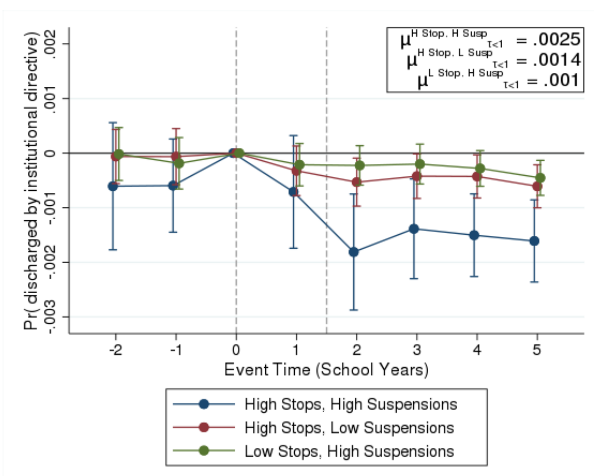
Notes: Figures display sensitivity analysis of estimated effects on felony and violent misdemeanor crimes (mean=49.52) to potential violations of the parallel trends assumptions per Rambachan and Roth (2021). The blue bar in each panel represents the 95% confidence interval of the difference-in-differences estimate for relative time τ from estimation of Equation 4 using precinct-clustered standard errors, except that we set the reference period to the year directly proceeding *Floyd* ($\tau = 0$). The red bars represent corresponding 95% confidence intervals when allowing for per-period violations of parallel trends of up to $\bar{M}$. That is, red confidence intervals impose that the change in the slope of the underlying trend be no more than M between consecutive periods.

Figure A.6: Effect on School Discharges by Changes to School Stop Exposure and Suspension Policy

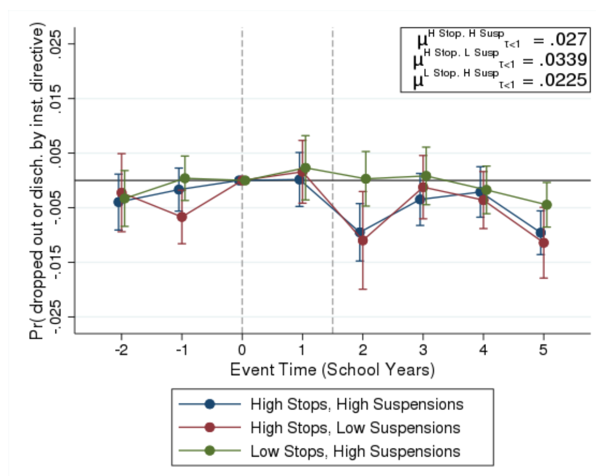
(a) Suspensions per High School Student



(b) *Pr(Discharged by Institutional Directive)*



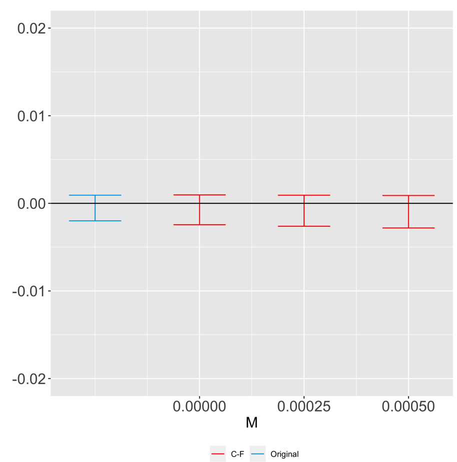
(c) *Pr(Dropped Out or Discharged by Inst. Dir.)*



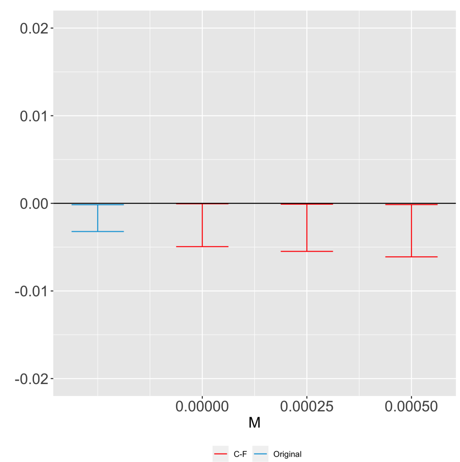
Notes: Panel A displays trends in total and level 2 "disorderly" suspensions for high and low suspension schools. High suspension schools are schools with above-median level 2 suspensions per student during the 2008 and 2009 school years. Panels B and C report coefficients from an adapted Equation 6 that collapses third and fourth quartile schools into a singular above-median stop-exposure indicator, and interacts this indicator with an above-median suspension indicator. The reference group is set to schools with below median stop-exposure and below-median suspensions. Whiskers mark 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the school-cohort level.

Figure A.7: Sensitivity Analysis for DD Estimates of *Floyd* on Institutional Discharge Rates

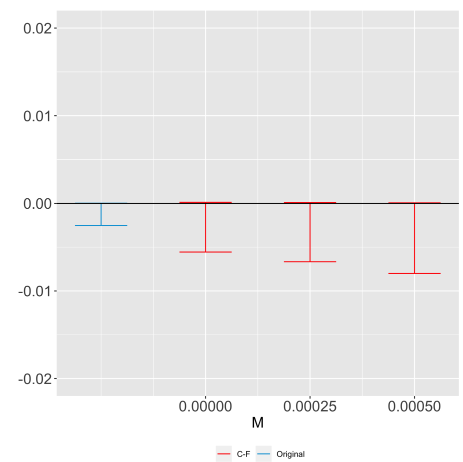
(a) $\tau = 1$



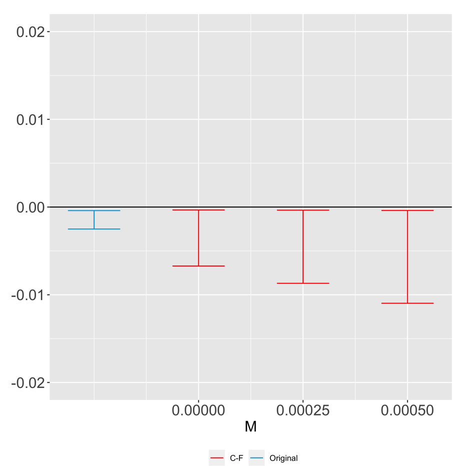
(b) $\tau = 2$



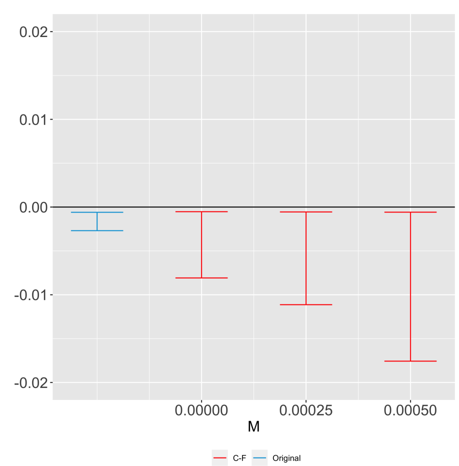
(c) $\tau = 3$



(d) $\tau = 4$



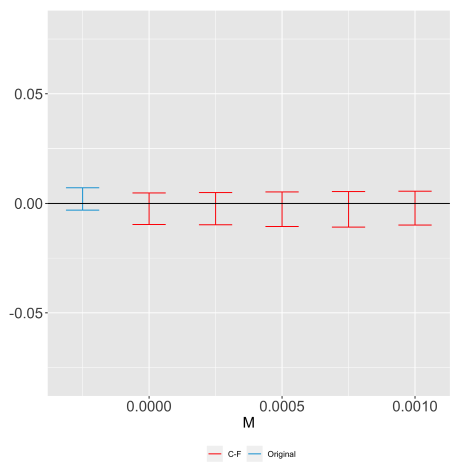
(e) $\tau = 5$



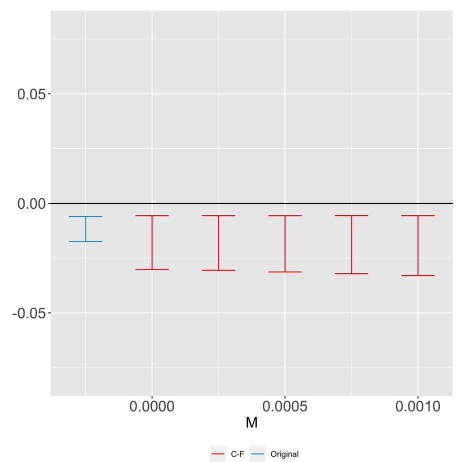
Notes: These figures display sensitivity analysis of estimated effects on the likelihood a student is dicharged by institutional directive. Specifically, we explore potential violations of the parallel trends assumptions per Rambachan and Roth (2021), assuming that the pre-trend differences. The blue bar in each panel represents the 95% confidence interval of the difference-in-differences estimate for relative time τ from estimation of Equation 6. Standard errors are clustered at the school-cohort level. Red bars correspond to 95% confidence intervals that allow for per-period violations of parallel trends of up to $\bar{M}$. That is, red confidence intervals impose that the change in the slope of the underlying linear trend be no more than M between consecutive periods.

Figure A.8: Sensitivity Analysis for DD Estimates of *Floyd* on Dropout Rates

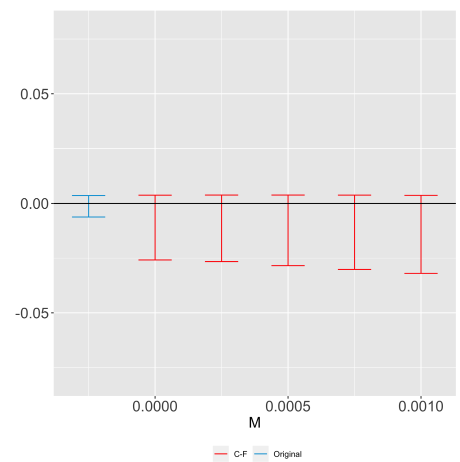
(a) $\tau = 1$



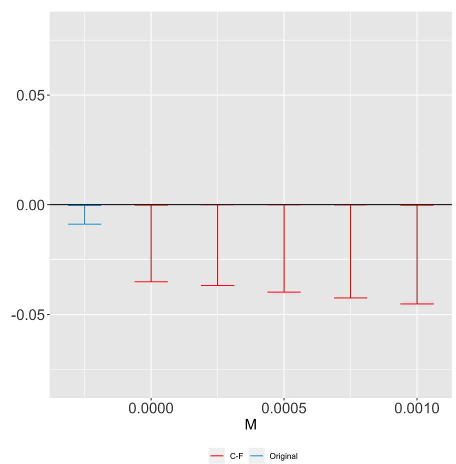
(b) $\tau = 2$



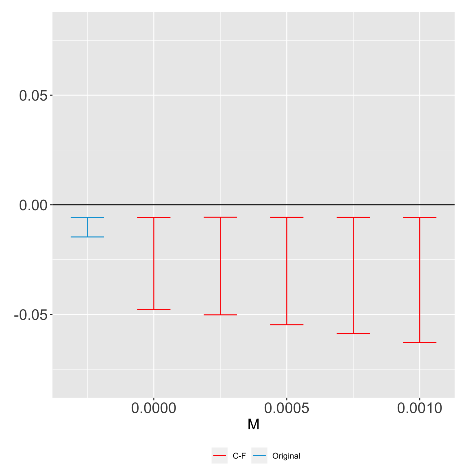
(c) $\tau = 3$



(d) $\tau = 4$



(e) $\tau = 5$

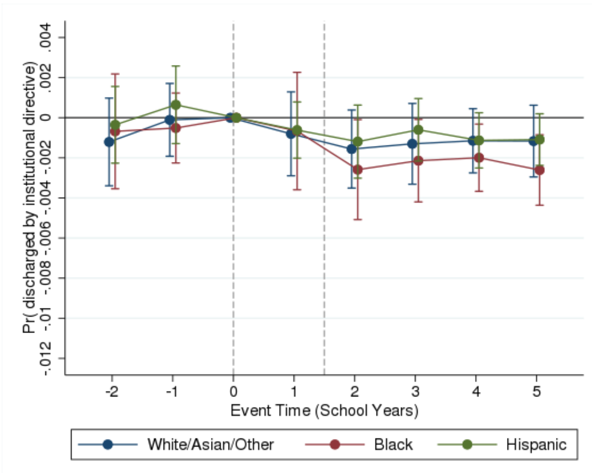


Notes: These figures display sensitivity analysis of estimated effects on the likelihood a student drops out or is dicharged by institutional directive. Specifically, we explore potential violations of the parallel trends assumptions per Rambachan and Roth (2021), assuming that the pre-trend differences lead to positive bias. The blue bar in each panel represents the 95% confidence interval of the difference-in-differences estimate for relative time τ from estimation of Equation 6. Standard errors are clustered at the school-cohort level. Red bars correspond to 95% confidence intervals that allow for per-period violations of parallel trends of up to $\bar{M}$. That is, red confidence intervals impose that the change in the slope of the underlying linear trend be no more than M between consecutive periods.

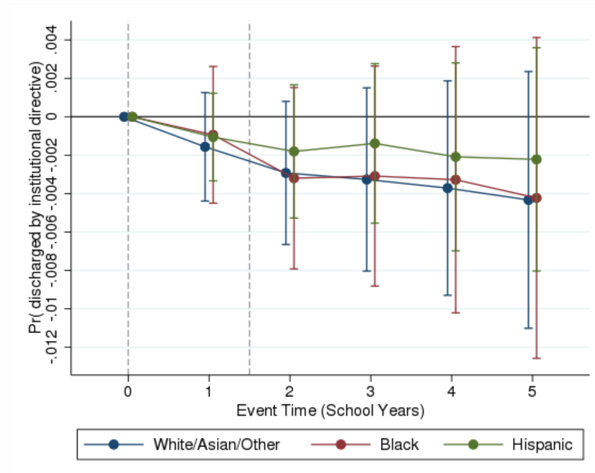
Figure A.9: Effect on Dropouts and Institutional Discharges by Student Race

DV = Pr(Discharged by Institutional Directive)

(a) Difference-in-differences

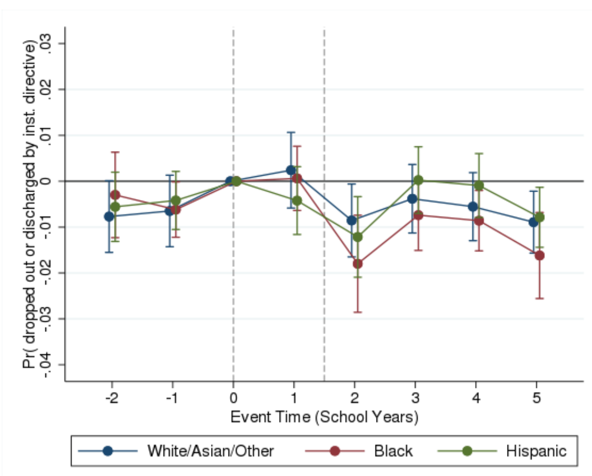


(b) + Linear Pre-trend Controls

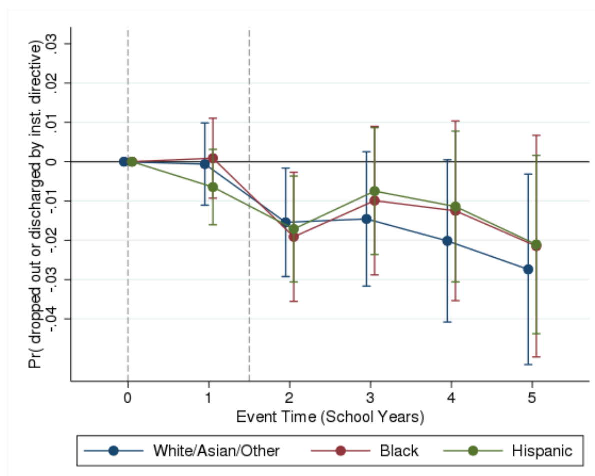


DV = Pr(Dropped Out or Discharged by Inst. Directive)

(c) Difference-in-differences



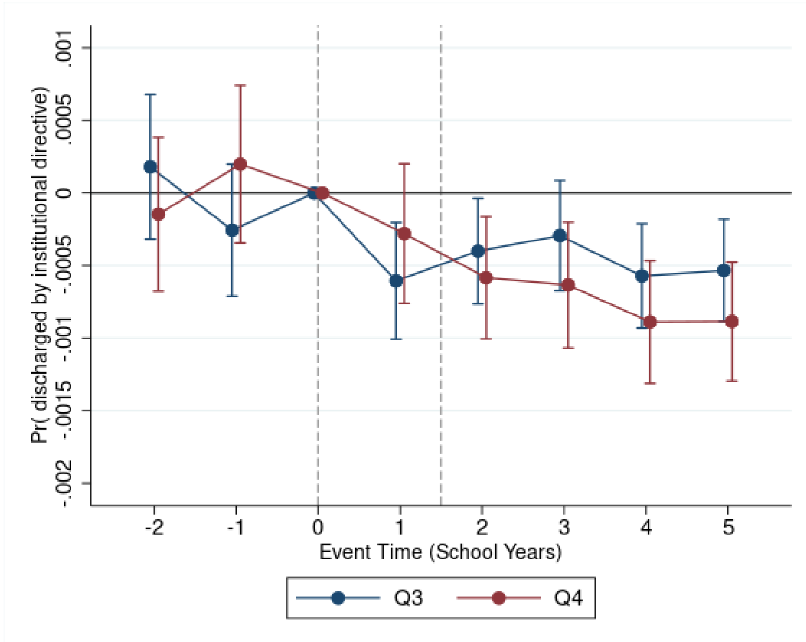
(d) + Linear Pre-trend Controls



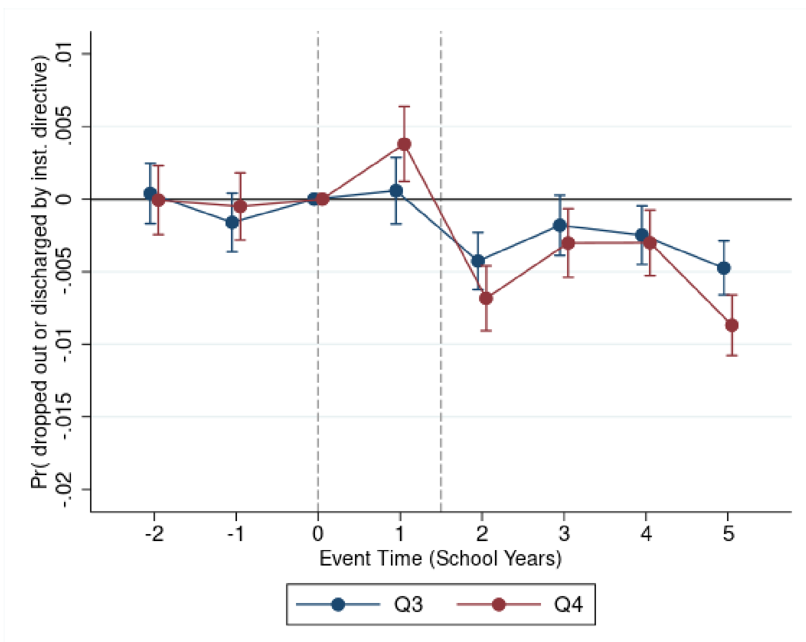
Notes: This figure graphs coefficients from Equation 6 in Panels A and C and coefficients from Equation 7 in Panels B and D, run separately for each racial sub-group. Only coefficients on Q4 time-to-treat indicators are included. Panels A and B display effects on the likelihood of enrolled students being discharged from high school by institutional directive during expected grades 9 through 12. Panels C and D display impacts on the likelihood of enrolled students dropping out or being discharged by institutional directive. The reference group is set to schools with below median exposure to stops during the training period. Estimates for white/Asian/other-race students, Black students, and Hispanic students are respectively reported in blue, maroon, and green. Whiskers mark 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the school-cohort level.

Figure A.10: Effect on High School Dropouts using Student-level Variation

(a) *DV = Pr(Discharged by Institutional Directive)*

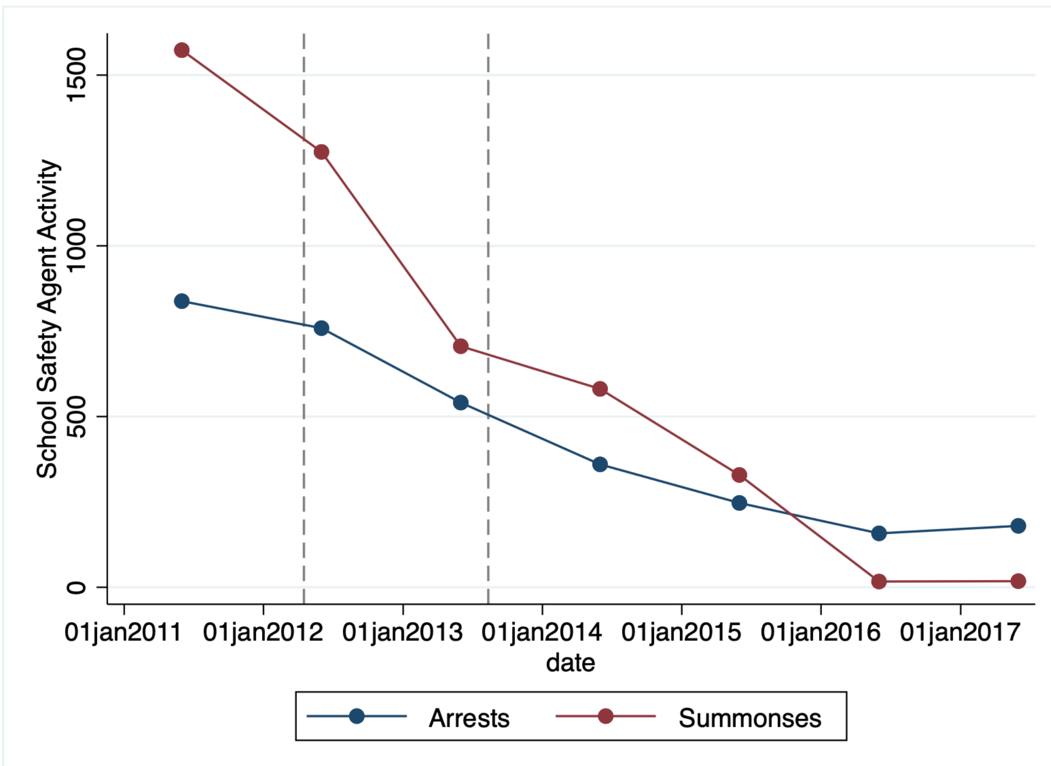


(b) *DV = Pr(Dropped Out or Discharged by Inst. Directive)*



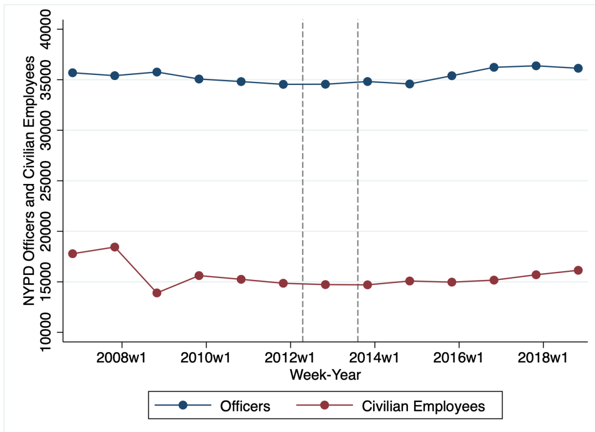
Notes: This figure graphs coefficients from Equation 10. Panel A displays effects on the likelihood of enrolled students being discharged from high school by institutional directive during expected grades 9 through 12. Panels B displays impacts on the likelihood of enrolled students dropping out or being discharged by institutional directive. The reference group is set to students from tracts with below median exposure to stops during the training period. Blue dots refer to point estimates for students living in third-quartile neighborhoods, while maroon does denote point estimates for students living in fourth-quartile or the "most exposed" neighborhoods. Whiskers mark 95% confidence intervals. Moving from left to right, the first vertical dashed line marks when police began scaling down stops in response to *Floyd*. The second dashed line marks the final ruling. All standard errors are clustered at the tract-by-cohort level.

Figure A.11: School Safety Agent Arrests and Court Summonses

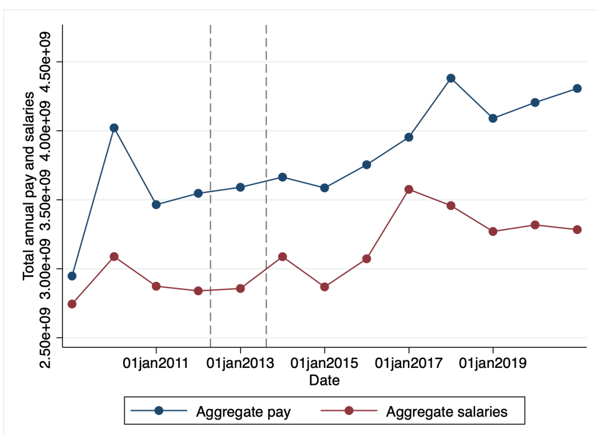


Notes: This table graphs the total number of arrests and court summonses issued by school safety officers in middle schools, high schools, and special education schools under the purview of the New York City Department of Education. Data come from the New York Civil Liberties Union and are available at `www.nyclu.org/en/student-safety-act-data`. Data are missing for the first quarter of 2011, so I impute 2011 totals by dividing the total number of stops observed in quarters two through four of 2011 by the $(1-q_1)$, where q_1 is the mean fraction of yearly arrests (summonses) observed in the first quarter during the other five years with complete data.

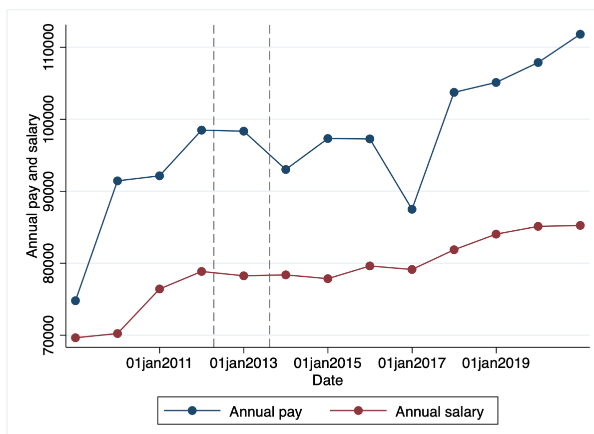
Figure A.12: NYPD Employment and Payroll Trends



(a) Total Pay Expenditures



(b) Pay per Officer



Notes: Panel A provides yearly counts of the number of officers employed by the NYPD. Data come from the FBI's Uniform Crime Reporting Program data. Specifically, we use data reported by the NYPD for the Law Enforcement Officers Killed and Assaulted (LEOKA) database for year 2006-2018. Data were collected from each agency as of October 31st of a given year. Data for Panels B and C are from NYC payroll records that are publicly available on `www.seethroughny.net/payroll`. We include all officers affiliated with the NYPD whose titles include "Police Officer, Special Officer, Sergeant, Lieutenant, Captain". This excludes school safety officers, school guards, traffic enforcement agents, police cadets, police chief, police commissioner, deputy commissioners and deputy chiefs.

D Supplemental Tables

Table A.1: Effect of *Floyd* on Neighborhood Crime in Highest Crime Neighborhoods

	Felony Rate Q3		Felony Rate Q4	
	Pre-per. T Mean	$\beta_{\tau\geq 1}$	Pre-per. T Mean	$\beta_{\tau\geq 1}$
Stops per 1mi^2/week	113.960	-50.906***	263.931	-115.148***
		(4.865)		(11.276)
Frisks per 1mi^2/week	66.026	-28.266***	144.071	-60.456***
		(3.267)		(6.703)
Uses of force per 1mi^2/week	24.623	-11.443***	58.087	-28.319***
		(2.055)		(5.010)
Stop-related arrests per 1mi^2/week	6.394	-2.420***	15.291	-5.869***
		(0.376)		(0.820)
Stop-related court summonses per 1mi^2/week	7.141	-3.258***	17.106	-6.663***
		(0.437)		(0.857)
Shootings per 1mi^2/week	0.279	-0.023	0.717	0.031
		(0.024)		(0.047)
Killings per 1mi^2/week	0.053	-0.003	0.121	0.003
		(0.010)		(0.014)
Felonies & violent misd per 1mi^2/week	26.512	-0.143	61.645	-1.003
		(0.375)		(0.910)
Major felonies per 1mi^2/week	12.997	0.103	29.308	-0.097
		(0.177)		(0.557)
Non-major fel & violent misd per 1mi^2/week	13.515	-0.246	32.337	-0.906
		(0.272)		(0.652)
Non-violent misd & violations per 1mi^2/week	32.766	-1.051*	87.033	-6.027***
		(0.608)		(2.255)

Notes: This table reports difference-in-differences estimates for various SQF and reported crime outcomes using Equation 3 with a singular post-*Floyd* indicator for each felony crime quartile as measured during the training period. "Pre-per. T mean" refers to the mean outcome in the Treatment group during the four years prior to *Floyd*. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Robustness of Estimated Effects of *Floyd* Neighborhood Crime

	Base	Alt. Time FE		Alt. Treat			Alt. Sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Stops per 1mi2/week								
Pre-per. Control Mean	51.471	51.471	51.471	61.336	32.523	59.855	58.695	75.920	107.477
Pre-per. Treat Mean	93.143	93.143	93.143	83.298	112.091	84.758	137.359	188.293	263.931
$\beta_{\tau\geq 1}$	-44.475***	-34.729***	-47.636***	-28.518**	-52.823***	-30.260***	-64.972***	-90.549***	-115.148***
	(7.031)	(8.937)	(8.962)	(12.240)	(5.735)	(6.003)	(7.769)	(8.608)	(11.276)
	Panel B: Felonies & violent misdemeanors per 1mi2/week								
Pre-per. Control Mean	27.457	27.457	27.457	25.623	21.866	27.783	31.368	40.588	57.888
Pre-per. Treat Mean	22.938	22.938	22.938	24.772	28.530	22.613	33.018	43.926	61.645
$\beta_{\tau\geq 1}$	-0.097	0.223	0.085	0.141	-0.510**	-0.487*	-0.204	-0.491	-1.003
	(0.229)	(0.268)	(0.264)	(0.518)	(0.232)	(0.276)	(0.317)	(0.494)	(0.910)
	Panel C: Major felonies per 1mi2/week								
Pre-per. Control Mean	14.406	14.406	14.406	13.327	12.029	14.173	16.413	21.055	29.589
Pre-per. Treat Mean	11.298	11.298	11.298	12.376	13.674	11.531	16.114	21.082	29.308
$\beta_{\tau\geq 1}$	0.123	0.001	0.090	0.121	-0.181	-0.153	0.094	0.032	-0.097
	(0.144)	(0.169)	(0.178)	(0.374)	(0.131)	(0.145)	(0.194)	(0.300)	(0.557)
	Panel D: Non-major felonies & violent misdemeanors per 1mi2/week								
Pre-per. Control Mean	13.052	13.052	13.052	12.296	9.836	13.610	14.955	19.533	28.299
Pre-per. Treat Mean	11.640	11.640	11.640	12.396	14.856	11.082	16.904	22.844	32.337
$\beta_{\tau\geq 1}$	-0.220	0.222	-0.005	0.020	-0.329*	-0.333*	-0.298	-0.523	-0.906
	(0.161)	(0.234)	(0.192)	(0.385)	(0.166)	(0.192)	(0.221)	(0.346)	(0.652)
Observations	1,009,944	1,009,944	1,009,944	1,009,944	1,009,944	1,009,944	757,224	504,972	252,252

Notes: This table shows DD coefficients from estimation of Equation 3, replacing event-time-treatment indicators with a post-treatment indicator. Panel A reports coefficients for stops per 1mi^2/week and Panel B explores felony and violent misdemeanor crimes per 1mi^2/week. Panels C and D disaggregate the outcome in Panel B into major felonies per 1mi^2/week and non-major felonies and violent misdemeanors per 1mi^2/week, respectively. Column 1 presents our base specification. Column 2 and 3 replace precinct-time FE with time FE and county-time FE, respectively. Column 4 assigns treatment based on precinct-level variation in stop rates during the training period by running Equation 2 at the precinct level. Since treatment varies at the precinct level, we replace precinct-time FE with county-time FE. Columns 5 and 6 explore alternative ways to define treatment at the tract level. Column 5 controls linearly for shootings, major felonies, non-major felonies, and misdemeanors in the training period. Column 6 adds various time-invariant neighborhood characteristics to Equation 2. These include indicators for having 0-10% or $\geq$ 10% of tract area covered by public housing or having, and decile fixed effects for the following neighborhood characteristics: % Black, % Hispanic, % families living below the poverty line, median household income, and % of adult residents with less than a HS degree. Columns 7-9 respectively restrict the sample to neighborhoods that are ranked in the top 75%, 50%, or 25% in terms of felony crimes per 1mi^2/week during the training period. All standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 2,058 Census tracts with 50-50 % split of Treatment and Control neighborhoods.

Table A.3: Effect of Impact Zone Assignment using Years from First Assignment Event

	Pre-period Mean	Year 1	Year 2	Year 3	P-value
Policing	(1)	(2)	(3)	(4)	(5)
Fraction of Tract Covered by Impact Zone	0.002	0.331***	0.247***	0.298***	0.000
		(0.033)	(0.034)	(0.047)	
Stops per $1mi^2$/week	140.360	38.004***	33.773**	45.674**	0.003
		(7.964)	(14.576)	(19.393)	
Officers with a Stop in 30-day Span	23.195	5.082***	4.666***	6.015***	0.000
		(0.984)	(1.467)	(2.098)	
Mean Stops per Officer	1.494	0.075***	0.036	0.051	0.109
		(0.025)	(0.033)	(0.049)	
Frisks per $1mi^2$/week	71.138	17.951***	22.867***	24.704***	0.001
		(4.724)	(7.906)	(9.047)	
Uses of Force per $1mi^2$/week	28.128	6.763***	8.879***	8.347*	0.009
		(1.977)	(3.292)	(4.857)	
Stop-related Arrests per $1mi^2$/week	6.736	0.035	0.172	-0.690	0.829
		(0.433)	(0.798)	(1.182)	
Stop-related Court Summonses per $1mi^2$/week	9.886	3.509***	4.534***	5.030***	0.002
		(0.850)	(1.708)	(1.851)	
Crime					
Felonies & Violent Misd per $1mi^2$/week	43.146	1.343	0.296	-1.004	0.878
		(0.990)	(1.453)	(1.867)	
Major Felonies per $1mi^2$/week	20.024	-0.871	-1.010	-1.661	0.277
		(0.760)	(1.077)	(1.463)	
Non-major Fel & Violent Misd per $1mi^2$/week	23.123	2.214***	1.306	0.657	0.081
		(0.540)	(0.900)	(1.116)	
Non-violent Misd & Violations per $1mi^2$/week	52.885	3.171***	2.584	2.346	0.124
		(1.136)	(1.812)	(2.608)	
Shootings per $1mi^2$/week	0.680	0.148*	0.238*	0.362*	0.071
		(0.086)	(0.136)	(0.206)	
Killings per $1mi^2$/week	0.134	0.025	0.018	0.023	0.661
		(0.037)	(0.052)	(0.070)	
Observations (tract-weeks)		65,964			
Census Tracts		303			

Notes: This table reports coefficients from an adapted version of Equation 5 that defines event-time (τ) as six-month periods relative to a Census tract's initial assignment to an Impact Zone from January 2006 through July 2012. We collapse event-time into yearly indicators to improve precision. "Pre-period mean" refers to the outcome mean in treated neighborhoods in the three years leading up to assignment. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Treatment is defined as $> 5\%$ of Census tract area covered by an Impact Zone.

Table A.4: Effect of Impact Zone Assignment on Neighborhood Policing and Crime by Distance from Impact Zone

	≤ 5% IZ Cov. & within 0.25 miles					0.25-1.0 miles from IZ				
	Pre-Mean	Yr 1	Yr 2	Yr 3	P-val	Pre-Mean	Yr 1	Yr 2	Yr 3	P-value
Policing	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fraction of Tract Covered by Impact Zone	0.000	-0.004	-0.005	-0.010	0.370	0.000	-0.003	-0.002	-0.008	0.049
		(0.004)	(0.009)	(0.011)			(0.002)	(0.003)	(0.004)	
Stops per 1mi^2/week	90.435	-1.779	-3.395	4.530	0.963	52.847	0.535	-6.868	-11.639	0.225
		(2.947)	(5.376)	(7.426)			(3.138)	(5.857)	(6.908)	
Officers with a Stop in 30-day Span	16.491	0.217	0.263	1.965	0.253	10.470	0.343	-0.399	-1.041	0.538
		(0.365)	(0.815)	(1.185)			(0.373)	(0.730)	(0.803)	
Mean Stops per Officer	1.482	-0.010	-0.021	-0.006	0.509	1.406	-0.004	0.008	-0.029	0.648
		(0.011)	(0.022)	(0.029)			(0.014)	(0.025)	(0.023)	
Frisks per 1mi^2/week	52.315	-0.225	-1.395	5.913	0.666	29.073	0.220	-5.078	-5.436*	0.169
		(1.651)	(4.209)	(5.340)			(1.527)	(3.519)	(3.195)	
Uses of Force per 1mi^2/week	21.953	1.813	3.948	6.438	0.131	11.968	-0.686	-5.240*	-3.467**	0.032
		(1.457)	(2.927)	(4.317)			(0.819)	(2.638)	(1.554)	
Stop-related Arrests per 1mi^2/week	5.402	0.352	0.091	0.499	0.398	2.964	-0.145	0.099	0.013	0.966
		(0.330)	(0.459)	(0.562)			(0.191)	(0.289)	(0.463)	
Stop-related Court Summ per 1mi^2/week	6.741	0.420	-0.138	0.786	0.462	3.802	0.141	0.082	0.197	0.767
		(0.352)	(0.535)	(0.874)			(0.323)	(0.563)	(0.713)	
Crime										
Felonies & Violent Misd per 1mi^2/week	36.099	0.056	1.240	0.834	0.158	22.219	0.243	0.083	-0.087	0.854
		(0.334)	(0.731)	(0.905)			(0.286)	(0.505)	(0.667)	
Major Felonies per 1mi^2/week	16.860	-0.374	0.019	0.343	0.989	11.681	0.105	0.087	0.235	0.602
		(0.255)	(0.432)	(0.501)			(0.152)	(0.317)	(0.483)	
Non-major Fel & Violent Misd per 1mi^2/week	19.240	0.431*	1.221**	0.491	0.056	10.538	0.138	-0.004	-0.322	0.807
		(0.217)	(0.527)	(0.631)			(0.194)	(0.314)	(0.387)	
Non-violent Misd & Violations per 1mi^2/week	47.393	1.004*	1.056	0.057	0.294	29.907	0.128	-0.263	-0.644	0.704
		(0.552)	(0.660)	(1.419)			(0.495)	(0.821)	(0.969)	
Shootings per 1mi^2/week	0.497	-0.029	-0.018	-0.046	0.765	0.236	-0.041	-0.025	-0.063	0.591
		(0.066)	(0.115)	(0.157)			(0.047)	(0.093)	(0.110)	
Killings per 1mi^2/week	0.096	-0.002	0.028	-0.008	0.875	0.049	-0.004	-0.012	-0.012	0.769
		(0.022)	(0.041)	(0.058)			(0.020)	(0.034)	(0.044)	
Observations (tract-weeks)			84,417					126,522		
Census Tracts			452					666		

Notes: This table reports coefficients from Equation 5 for various policing and crime measures. We collapse event-time into yearly indicators to improve precision. "Pre-per. mean" refers to the outcome mean in treated neighborhoods in the three years leading up to assignment. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. We explore two separate definitions of treatment: tracts with ≤ 5% IZ coverage or within 0.25-miles of an active Impact Zone, and tracts within 0.25 to 1.00 miles from an active Impact Zone. SEs are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Effect of Impact Zone Assignment on Neighborhood Crime for Tracts within 0.25 Miles of Impact Zone

	Pre-period Mean	Year 1	Year 2	Year 3	P-value
Policing	(1)	(2)	(3)	(4)	(5)
Fraction of Tract Covered by Impact Zone	0.001	0.116***	0.121***	0.114***	0.000
		(0.014)	(0.019)	(0.030)	
Stops per 1mi^2/week	113.577	16.826***	22.912**	43.896***	0.000
		(4.122)	(9.585)	(12.137)	
Officers with a Stop in 30-day Span	19.350	2.041***	2.850***	5.108***	0.000
		(0.463)	(1.069)	(1.425)	
Mean Stops per Officer	1.498	0.015	0.013	0.035	0.194
		(0.011)	(0.016)	(0.029)	
Frisks per 1mi^2/week	62.707	8.327***	16.127**	21.536***	0.002
		(2.841)	(6.144)	(7.540)	
Uses of Force per 1mi^2/week	25.826	4.919**	11.147**	8.825**	0.028
		(2.329)	(5.091)	(4.426)	
Stop-related Arrests per 1mi^2/week	6.213	0.478*	0.114	0.640	0.153
		(0.267)	(0.388)	(0.429)	
Stop-related Court Summonses per 1mi^2/week	8.168	1.298**	2.399***	3.605***	0.000
		(0.532)	(0.872)	(1.287)	
Crime					
Felonies & Violent Misd per 1mi^2/week	39.960	0.314	0.259	-0.324	0.877
		(0.325)	(0.700)	(0.780)	
Major Felonies per 1mi^2/week	18.360	-0.675***	-0.832**	-0.899*	0.017
		(0.239)	(0.408)	(0.470)	
Non-major Fel & Violent Misd per 1mi^2/week	21.600	0.989***	1.090**	0.575	0.030
		(0.269)	(0.521)	(0.606)	
Non-violent Misd & Violations per 1mi^2/week	51.990	1.276**	0.992	0.165	0.456
		(0.627)	(1.233)	(1.586)	
Shootings per 1mi^2/week	0.584	0.045	0.066	0.119	0.368
		(0.053)	(0.091)	(0.129)	
Killings per 1mi^2/week	0.111	0.006	0.050	0.071	0.234
		(0.020)	(0.037)	(0.054)	
Observations (tract-weeks)			149,001		
Census Tracts			640		

Notes: This table reports coefficients from Equation 5 for various policing and crime measures. We collapse event-time into yearly indicators to improve precision. "Pre-per. mean" refers to the outcome mean in treated neighborhoods in the three years leading up to assignment. Standard errors are clustered at the precinct-level and reported in parentheses below coefficients. We define treatment as any Census tract within 0.25-miles of an active Impact Zone (including those covered by Impact Zones). SEs are clustered at the precinct-level and reported in parentheses below coefficients. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Effect of *Floyd* on Dropout and Institutional Discharge Rates by Pre-Reform School Suspension Rates

	Pre-period	Base		+ Linear Pre-trends	
	Treat Mean	$\beta_{\tau \geq 2}$	P-value	$E[\beta_{\tau \in [2,5]}]$	P-value
	(1)	(2)	(3)	(4)	(5)
Pr(Discharged by Institutional Directive)					
High Stops, High Suspensions	0.00252	-0.00118***	0.00049	-0.00254*	0.06175
		(0.00034)			
High Stops, Low Suspensions	0.00140	-0.00045***	0.00064	-0.00060	0.30049
		(0.00013)			
LowStops, High Suspensions	0.00096	-0.00022*	0.09561	-0.00026	0.63818
		(0.00013)			
Pr(Dropped Out or Discharged by Inst. Directive)					
High Stops, High Suspensions	0.02699	-0.00435***	0.00281	-0.01321**	0.02724
		(0.00145)			
High Stops, Low Suspensions	0.03386	-0.00386*	0.08457	-0.00872	0.26514
		(0.00224)			
LowStops, High Suspensions	0.02245	-0.00027	0.85216	-0.00788	0.18591
		(0.00142)			

Notes: Adapting Equations 6 and 7, this table reports post-ruling coefficients on interactions between an above-median stop-exposure indicator with above-median suspension exposure indicator. I report DD estimates for [High Stops, High Suspension] schools, [High Stops, Low Suspension] schools, and [Low Stops, High Suspension] schools relative to [Low Stops, Low Suspension] schools. Both indicators are measured at the school-level. Above-median suspensions are measured as mean Level 2 "disorderly" suspensions per high school student enrolled from 2008 to 2009 school years. Column 1 reports mean outcomes during 2010 to 2012 school years. Column 2 reports post-ruling impacts (i.e. years 2014 through 2017). Standard errors are reported below coefficients in parentheses. Column 4 reports the average across all four post-ruling coefficients, and Column 5 reports the p-value from an F-test that the four post-ruling coefficients are jointly zero. Standard errors are clustered at the school-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Effect of *Floyd* on Dropout Rates using Student-level Variation

	Pre-period	Base		+ Linear Pre-trends	
	Treat Mean	$\beta_{\tau\geq 2}$	P-value	$E[\beta_{\tau\in[2,5]}]$	P-value
	(1)	(2)	(3)	(4)	(5)
DV = Stops per year					
Third Quartile	408.6499	-257.6284***	0.000	-245.6365***	0.000
		7.4320			
Fourth Quartile	923.8398	-763.1276***	0.000	-829.749***	0.000
		12.2966			
DV = Stops of 14-18 year-old residents per year					
Third Quartile	86.1784	-53.1894***	0.000	-58.9141***	0.000
		1.7777			
Fourth Quartile	194.2284	-159.5676***	0.000	-211.1002***	0.000
		3.4112			
DV = Pr(Discharged by Institutional Directive)					
Third Quartile	0.0015	-0.0004***	0.000	0.0000	0.970
		0.0001			
Fourth Quartile	0.0021	-0.0008***	0.000	-0.0011*	0.072
		0.0001			
DV = Pr(Dropped Out or Discharged by Inst. Directive)					
Third Quartile	0.0267	-0.0029***	0.000	-0.0020	0.403
		0.0005			
Fourth Quartile	0.0322	-0.0052***	0.000	-0.0054*	0.055
		0.0006			

Notes: This table reports coefficients from Equation 10 and an adapted version that controls for linear differences in pre-trends. Treatment is defined as students from neighborhoods with stop exposure during the training period that ranked between the 50th and 75th percentile – "third quartile" – or 75-100th percentile – fourth quartile. Column 1 reports mean outcomes for each quartile during the 2010 through 2012 school years. Column 2 reports the coefficient on a post-ruling indicator that collapses 2014-2017 treatment indicators in Equation 10 into a singular indicator. Standard errors are reported below coefficients in parentheses. Column 4 reports the average across all four post-ruling coefficients controlling for linear pre-trend differences, and Column 5 reports the p-value from an F-test that the four post-ruling coefficients are jointly zero. Standard errors are clustered at the tract-by-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Effect of *Floyd* on Institutional Discharge Rates by Student Race and Sex

Y = Pr(Discharged by Institutional Directive)	Pre-period Treat Mean	Base $\beta_{\tau\geq 2}$	Base P-value	+ Linear Pre-trends $E[\beta_{\tau\in[2,5]}]$	+ Linear Pre-trends P-value
	(1)	(2)	(3)	(4)	(5)
Student Race					
Black	0.0036	-0.0019**	0.0155	-0.0034	0.2869
		(0.0008)			
Hispanic	0.0018	-0.0011*	0.0619	-0.0019	0.4177
		(0.0006)			
White/Asian/Other	0.0020	-0.0009	0.1235	-0.0036	0.1729
		(0.0006)			
Student Sex					
Male	0.0042	-0.0021**	0.0101	-0.0035	0.3354
		(0.0008)			
Female	0.0010	-0.0007***	0.0020	-0.0011	0.3549
		(0.0002)			
Student Race x Sex					
Black Male	0.0059	-0.0027*	0.0537	-0.0054	0.4294
		(0.0014)			
Black Female	0.0013	-0.0011**	0.0020	-0.0019	0.2979
		(0.0004)			
Hispanic Male	0.0028	-0.0018*	0.0819	-0.0040	0.2773
		(0.0010)			
Hispanic Female	0.0008	-0.0005	0.1133	0.0000	0.9989
		(0.0003)			
White/Asian/Other Male	0.0029	-0.0009	0.2670	-0.0038	0.3836
		(0.0008)			
White/Asian/Other Female	0.0009	-0.0007	0.2345	-0.0027	0.2065
		(0.0006)			

Notes: This table reports post-ruling coefficients for fourth quartile schools from Equations 6 and 7 ran separately restricting the data to each sub-group. The outcome of interest is an indicator for whether the student switched from being enrolled at the beginning of the school year to being discharged by institutional directive. Column 1 reports mean outcomes for each sub-group group during the 2010 through 2012 school years. Column 2 reports race-specific post-ruling impacts (i.e. years 2014 through 2017) using Equation 6. Standard errors are reported below coefficients in parentheses. Column 4 reports the average across all four post-ruling coefficients from Equation 7, and Column 5 reports the p-value from an F-test that the four post-ruling coefficients are jointly zero. Standard errors are clustered at the school-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Effect of *Floyd* on Dropout Rates by Student Race and Sex

Y = Pr(Dropped Out or Discharged by Inst. Dir.)	Pre-period Treat Mean	Base		+ Linear Pre-trends	
		$\beta_{\tau \geq 2}$	P-value	$E[\beta_{\tau \in [2,5]}]$	P-value
	(1)	(2)	(3)	(4)	(5)
Student Race					
Black	0.0290	-0.0095***	0.0007	-0.0157	0.1338
		(0.0028)			
Hispanic	0.0297	-0.0021	0.3664	-0.0143	0.1002
		(0.0023)			
White/Asian/Other	0.0237	-0.0021	0.2919	-0.0194**	0.0397
		(0.0020)			
Student Sex					
Male	0.0326	-0.0035*	0.0672	-0.0196**	0.0143
		(0.0019)			
Female	0.0253	-0.0036**	0.0133	-0.0139**	0.0222
		(0.0015)			
Student Race x Sex					
Black Male	0.0335	-0.0114***	0.0016	-0.0161	0.2608
		(0.0036)			
Black Female	0.0247	-0.0081***	0.0021	-0.0149	0.1289
		(0.0026)			
Hispanic Male	0.0329	-0.0022	0.4513	-0.0221*	0.0518
		(0.0029)			
Hispanic Female	0.0265	-0.0012	0.5972	-0.0075	0.4182
		(0.0023)			
White/Asian/Other Male	0.0265	-0.0021	0.4291	-0.0205	0.1111
		(0.0027)			
White/Asian/Other Female	0.0203	-0.0017	0.4998	-0.0189	0.1084
		(0.0025)			

Notes: This table reports post-ruling coefficients for fourth quartile schools from Equations 6 and 7 ran separately restricting the data to each sub-group. The outcome of interest is an indicator for whether the student switched from being enrolled at the beginning of the school year to being listed as dropped out or discharged by institutional directive. Column 1 reports mean outcomes for each sub-group group during the 2010 through 2012 school years. Column 2 reports race-specific post-ruling impacts (i.e. years 2014 through 2017) using Equation 6. Standard errors are reported below coefficients in parentheses. Column 4 reports the average across all four post-ruling coefficients from Equation 7, and Column 5 reports the p-value from an F-test that the four post-ruling coefficients are jointly zero. Standard errors are clustered at the school-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.