Essays on the Economics of Policing and Crime

Roman Gabriel Rivera

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

Essays on the Economics of Policing and Crime

Roman Gabriel Rivera

There is growing demand for reforms to the U.S. criminal justice system. Nevertheless, there are significant questions and relatively few answers. This dissertation studies multiple U.S. criminal justice system issues using detailed administrative data from Cook County, Illinois: Does policing the police increase crime? Does the composition of a police officer's academy cohort influence their future outcomes? Is pretrial electronic monitoring an attractive alternative to pretrial release and detention? To answer these questions, I use administrative data from Chicago and Cook County, Illinois, on the Chicago Police Department, Cook County Jail, and Circuit Court of Cook County, and a range of econometric methods.

In Chapter 1, I study the effect of pretrial electronic monitoring (EM) as an alternative to pretrial release and pretrial detention (jail) in Cook County, Illinois. EM often involves a defendant wearing an electronic ankle bracelet that tracks their movement and aims to deter pretrial misconduct. Using the quasi-random assignment of bond court judges, I estimate the effect of EM versus release and EM versus detention on pretrial misconduct, case outcomes, and future recidivism. I develop a novel method for the semiparametric estimation of marginal treatment effects in ordered choice environments, with which I construct relevant treatment effects. Relative to release, EM increases new cases pretrial due to bond violations while reducing new cases for low-level crimes and failures to appear in court. Relative to detention, EM increases low-level

pretrial misconduct but improves defendant case outcomes and reduces cost-weighted future recidivism. Finally, I bound EM's pretrial crime reduction effect. I find that EM is likely an adequate substitute for pretrial detention. However, it is unclear that EM prevents enough high-cost crime to justify its use relative to release, particularly for defendants who are more likely to be released.

Chapter 3, joint with Bocar Ba, studies and differentiates between the effects of oversight and outrage on policing. Previous studies estimating the impact of police oversight on crime rely on major policing scandals as shocks to examine the impact of oversight on crime. We argue that the simultaneous effect of public outrage on officer behavior and crime contaminates these results, and we provide a conceptual framework that distinguishes between oversight and outrage. We identify two events relating to unexpected court rulings in Chicago that increased oversight and caused a decline in reported misconduct but had virtually no public reaction. Despite the decrease in reported misconduct, crime and officer activity were unaffected. We contrast this with a major policing scandal, after which we find both a rise in crime rates without an equivalent increase in arrests and a decline in officer stops and use of force. Our results suggest that police oversight can reduce misconduct without increasing crime.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am forever grateful to my wife, Amani Abou Harb, for her endless patience and understanding and for her tolerance for reading my many drafts over many years, and for our daughter, Sophia. I thank my family for their support during my Ph.D.: my parents, José and Margret-Anne; my siblings, Catherine, Michael, Rakesh, and Vivek, and my siblings-in-law, Ali, Farah, and Mohammad; my step-mother, Tejal; my abuelos and grandparents, José and Gladys, Jim and Gail, and Usha; and my mother-in-law, Iman. I thank Bocar Ba for almost a decade of collaboration and friendship and for introducing me to the world of research.

I am indebted to my dissertation committee, Sandra Black, W. Bentley MacLeod, Simon Lee, Brendan O'Flaherty, and Bocar Ba, for their comments and feedback on this work. I thank Sandra Black and W. Bentley MacLeod for their dedication to my academic and personal success during my Ph.D. and for many hours of feedback on multiple papers. I also thank Simon Lee for his insights and commitment to developing my job market paper.

I am forever indebted to Robert LaLonde, whose early belief in my potential and precious time made all the difference.

I am grateful to my first-year study group, Wendy Morrison, Mitchell Vaughn, and Avery Dao, and to Kate Daloz for allowing us to use the GSAS Writing Studio long before we were actually writing our dissertations.

This work would not be possible without the generosity of the Chicago Data Collaborative, Invisible Institute, and Rachel Ryley, who shared crucial data for this and many other projects, or without Ali Ammoura and his guidance and insight into the court system of Cook County.

I thank my many coauthors, from whom I have learned much: Bocar Ba, Justin Holz, Nayoung Rim, Andrea Kiss, Jonathan Mummolo, Dean Knox, Jacob Kaplan, Pat Bayer, Rachel Mariman, Modibo Sidibe, Mayya Komisarchik, and Michelle Torres.

I thank the many researchers who have provided me with comments and feedback on my work and support in my career, including: Nour Abdul-Razzak, Amani Abou Harb, Amanda Agan, Livia Alfonsi, Douglas Almond, Francisca Antman, Bocar Ba, Jason Baron, Pat Bayer, Michael Best, Tarikua Erda, Felipe Goncalves, Bhargav Gopal, Abby Grobbel, Sakshi Gupta, Michelle Jiang, Jenny Jiao, Taeho Kim, Dean Knox, Emily Leslie, Steve Mello, Annie McGowan, Claire Montialoux, Jonathan Mummolo, Brendan O'Flaherty, José Luis Montiel Olea, Aurélie Ouss, Nayoung Rim, Bernard Salanié, Rajiv Sethi, Ashley Swanson, Miguel Urquiola, Emily Weisburst, and Caitlin Yee.

I thank Amy Claessens, Nicholas Polson, and Koichiro Ito for guidance in my Ph.D. applications and Erik Hurst for sponsoring my visit to Chicago Booth.

I am incredibly grateful to the members of the Invisible Institute and others whose work and collaboration provided me with the knowledge and data for much of my research, particularly Chaclyn Hunt, Emma Herman, Sam Stecklow, and Andrew Fan. I thank Patrick Ball for the most valuable coding instruction I've ever had and Kathie Kane-Willis for introducing me to the importance of policy research.

I thank the Ford Foundation Fellowship Program and the National Academies of Sciences, Engineering, and Medicine, and the Program for Economic Research for their generous financial support.

Dedication

To Amani and Sophia.

Chapter 1: Release, Detain, or Surveil?

Roman Gabriel Rivera

1.1 Introduction

The social and economic costs of pretrial detention are massive: around half a million individuals are detained while presumed innocent on any given day in the United States, resulting in a direct cost to local governments of around $14 billion annually.[1] However, releasing defendants pretrial risks increased crime. Over the last two decades, an alternative to both pretrial detention and release, electronic monitoring (EM) — technology that surveils and limits the movement of defendants using an ankle bracelet — has expanded across the United States, and its adoption has accelerated due to the COVID-19 pandemic.[2] Proponents of EM believe it promises better outcomes for defendants and lower costs to taxpayers than pretrial detention ([11]) while avoiding the potential increase in crime if defendants were released from jail without EM. Critics of EM argue that its expansion leads to unnecessary surveillance and additional criminal charges against defendants for non-criminal violations while increasing their contact with the criminal justice system ([4]).[3]

Two questions vital to understanding the value of EM remain unanswered in the literature. First, relative to detention, is the cost of pretrial crime under EM outweighed by EM's lower direct costs to taxpayers and better outcomes for defendants? Second, relative to release, does EM prevent enough pretrial misconduct to outweigh its adverse effects on defendants, such as its increase in rearrests due to EM violations? Both of these questions hinge upon how sensitive

[1] See [1] and [2].

[2] In 2015, more than 125,000 people in the US were on EM, more than double the number in 2005 ([3]). [4] reviews EM's popularity across the US. [5], [6], [7], [8], [9], and [10] discuss the expansion of the use of EM in specific US municipalities.

[3] For additional discussion of the debate around EM and criticisms, see [12], [13], [14], [15], [16], [17], and [18].

defendant decision-making is to surveillance and how many more rearrests occur on EM due to non-criminal violations.

The primary reason for the lack of evidence on the effect of pretrial EM relative to release and detention is a lack of data. Furthermore, because EM is a substitute for both release and detention — two very different alternatives — understanding its effects requires differentiating between both margins. In this paper, I use administrative court and jail data from Cook County, IL, one of the largest bond courts in the US and an early mass-adopter of EM. I document the effects of pretrial EM relative to both release and detention on pretrial misconduct, case outcomes, and future recidivism. I leverage the quasi-random assignment of bond court judges to recover the effects of EM versus release and EM versus detention using two-stage least squares (2SLS). While 2SLS can recover the average treatment effects on compliers, the average effect if EM were expanded to a larger population is more relevant for policy. I explore heterogeneity in treatment effects on both margins using a marginal treatment effects (MTE) framework and a novel semiparametric estimation method. Then, I construct treatment effects that are relevant for policy, such as the average effect of expanding EM to the defendants who were released or to those who were detained.

Relative to detention, I find that EM allows for increased pretrial misconduct but overall improves defendant outcomes. Defendants placed on EM are more likely to fail to appear in court and to have a new case opened against them pretrial; low-level criminal charges and charges for bond violations drive the increase in new cases. However, I find suggestive evidence that EM is less coercive and criminogenic than detention, as it lowers the likelihood of incarceration and either decreases or has a null effect on cost-weighted post-trial recidivism.[4] EM does not increase pretrial crime relative to detention when new charges are weighted by the dollar cost of crime, making it a reasonable alternative to detention for most defendants. However, selection patterns indicate that the benefits of EM are smaller for defendants who are more likely to be detained.

For the effect of EM relative to release, EM reduces failures to appear in court, has no effect on new cases pretrial with serious charges, and weakly decreases new cases pretrial with low-level

[4] I follow [19] in my construction of a cost-weighted crime index.

charges for most defendants. The reduction in pretrial new cases with crime-related charges may understate EM's crime-reducing effect because EM increases the probability of detection through surveillance. To account for the change in detection rates, I compute the implied dollar cost of crime prevented by EM relative to release. Under very generous assumptions, the amount is less than \$10,000 for the average defendant and is smaller for released defendants, though I cannot reject a small or null effect.[5] Furthermore, I bound defendants' elasticities of pretrial crime with respect to the probability of detention. I find evidence consistent with defendants being elastic for low-level crimes but relatively inelastic for serious crimes.

EM imposes costs on defendants relative to release that may outweigh its potential benefits. EM increases the likelihood of a new case pretrial with charges related to bond violations and "escapes" because of the increased restrictions of movement while on EM. Naturally, these criminal charges for violations are socially costly because their punishment (e.g., re-incarceration and conviction) outweighs the harm of these activities. This highlights the trade-off with surveillance systems used to prevent undesirable behavior: additional conditions placed on individuals to enforce compliance with the system can lead to negative effects (e.g., charges for bond violation) that may undermine the benefits of reduced misconduct (e.g., crimes).

Furthermore, while the effects of EM relative to release on case outcomes and post-trial recidivism are mixed, the main results suggest that EM leads to weakly worse case outcomes and more socially costly future recidivism for most defendants. In general, selection patterns indicate EM's costs and benefits relative to release are both smaller for released defendants than for the average defendant. In contrast with EM's sizable benefits over detention, it is not clear that EM's benefits over release outweigh its costs.

This paper makes the following contributions to our understanding of the economics of pretrial detention and surveillance. First, existing work on EM in economics studies non-US contexts and has focused on the effect of EM relative to detention, pre- or post-trial, with recidivism being the

[5] Specifically, assuming EM detects 100% of crimes, all guilty charges are the crimes actually committed, and the crime-cost weighted probability of detection on release is 50%.

main outcome of interest ([20], [21], [22]).[6] This paper advances our understanding of how EM influences defendant outcomes by applying a new methodology to recover heterogeneous treatment effects of EM relative to both release and detention on pretrial misconduct, case outcomes, and recidivism in a large US municipality. Beyond the location and multitude of outcomes, these results better inform policy by allowing us to explore the effects of EM on defendants who are more or less likely to be released or detained — i.e., the effect of expanding EM on either margin. Relatedly, this paper builds on the pretrial detention and judge-design literatures by incorporating multiple pretrial treatments, while prior work generally focuses on a binary set of pretrial treatments (e.g., release versus detain), as well as the literatures on probation and alternatives to detention.[7]

This paper builds on the large literature on agency theory, in which principals use costly monitoring mechanisms to deter agents from 'shirking' and to improve productivity ([55], [56]).[8] I find that EM imposes a cost by increasing new cases pretrial due to non-criminal violations (relative to both release and detention). This cost is a byproduct of EM's inability to perfectly detect criminal activity, thereby punishing a wider set of (non-criminal) activities. This highlights the importance of considering not only the direct costs of the monitoring technology (in this case, maintaining the EM system) but also the cost of punishing violations of the monitoring system itself.[9] Consistent with models of agency theory and specific deterrence ([71]), I also find evidence consistent with EM reducing pretrial crime by increasing the probability of detection.[10] Furthermore, this paper connects agency theory to the literature on modern surveillance technology by studying the rise of individualized surveillance used to deter violations, such as body-worn cameras on police used to

[6]Additional work on EM includes [23], [24], [25], and [26]. See [27] for a review of the interdisciplinary literature on EM.

[7]See [28], [29], [30], [31], [32], [33], and [34] for 'judge' designs in pretrial context. Exploiting 'judge' random assignment is a common identification strategy: [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], and [46]. See [47] for a discussion of the US pretrial system. See [48] for an early analysis of the economics of pretrial detention. For work on the effect of probation, parole, and alternatives to detention, see [49], [50], [51], [52], [53], and [54].

[8]See also [57], [58], [59], [60], [61], and [62].

[9]This is consistent with criticisms of EM and surveillance in the criminal justice system more broadly ([18]) and highlights the importance of designing policy to minimize crime rather than maximize arrests ([63]). This also connects to the literature on elasticities of crime with respect to sentencing ([64], [65]) and detection rates ([66], [67], [68], [69], [70]).

[10]The probability of detection plays a major role in crime reduction through deterrence. See [72], [73], [74], [75], [76], and [77].

deter misconduct ([78]) or remote-work monitoring technologies ([79]), whereas much prior work focuses on the economics of mass surveillance technologies ([80]).[11]

Methodologically, this paper contributes to the literature on marginal treatment effects and identification with multiple treatments ([85], [86], [87]).[12] Much like [101], I build on [102] by showing that identification of marginal treatment response (MTR) functions in ordered treatment environments can be achieved relying solely on variation in the probability of adjacent treatments.[13] I develop a straightforward method for the semiparametric estimation of MTRs in ordered treatment environments, which can be applied in a range of ordered treatment environments and allows for non-monotonic MTEs in contrast with the existing fully parametric method for estimating MTEs with ordered treatments ([102], [104]).

This article proceeds as follows. Section 1.2 discusses the institutional background, data, and the potential costs and benefits of each treatment. Section 1.3 presents the empirical strategy and results for the 2SLS analysis. Section 1.4 presents the empirical strategy and results for the MTE analysis. Section 1.5 discusses policy implications, and Section 1.6 presents robustness checks. Section 1.7 concludes.

1.2 Background and Data

1.2.1 Cook County Bond Court, Bond Types, and Treatments

Bond Court Following an arrest in Cook County, a defendant is taken into custody and "booked" based on their arrest charges, generally at the Cook County Jail (CCJ). Then they are arraigned at bond court, usually within 1 day of the arrest. For the entire "pretrial" period, the defendant is presumed innocent.

This paper focuses on the central bond court in Cook County, known as "Branch 1". Branch 1 handles almost all felony cases in Cook County and operates every day. On weekdays and non-

[11] See also [81], [82], and [83] on AI and mass surveillance. [84] discusses the post-COVID rise of surveillance and remote-work.

[12] See also [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], and [98]. See also [99] and [100].

[13] [101] develops a similar method, but their focus is on bounding MTRs (building on [103]) rather than point-identification and estimation of MTRs over common support using continuous instruments.

holidays, non-felony cases (e.g., misdemeanor, traffic, and municipal code violations that require the setting of bail) are generally handled in the bond courts determined by the location of arrest (Branch 1 is the bond court for Chicago arrests) or specifically designated courts — for example, murder and violent sex offenses are handled in their own courts. On holidays and weekends, however, all such cases are handled by Branch 1. At bond court, the sitting bond court judge determines the bond conditions for a defendant, namely bond type and amount. [14]

During the period of this study, from July 2013 through 2015, a single bond court judge handled the cases that passed through Branch 1 on any given day. The judges have an irregular working schedule ([105]), which depends on their off days, vacations, and work-day preferences — Figure 1.1 displays a sample of the calendar from the data. There are relatively few active bond court judges within a given month (≈ 4), and only 7 were active during the period of study.[15] Defendants do not have discretion over their assigned judge, and judges cannot choose the cases they see on a given day. Importantly, because Branch 1 is always active (including holidays and weekends) and the schedule of each bond court judge is sporadic, there is no scheduling relationship between bond court judges and prosecutors, public defenders, or trial judges.

Bond Types In the period of study, there were three main bond types used by Branch 1 judges: D-bonds, I-bonds, and IEM bonds. A D-bond is the most common bond (55% of sample bonds) and conforms to the popular understanding of bonds: it requires a defendant to post 10% of the bond amount in order to be released from jail. This 10% amount can range widely, from below \$50 to over \$200,000. The defendant can pay the amount at any point during their pretrial period and be released from jail, or they are detained in jail until the pay or their case concludes.

I-bonds (15%), or "release on recognizance" bonds, do not require the defendant to post any money in order to be released from jail. However, as with all bond types discussed, the defendant is liable to pay the full bond amount if they violate their bond conditions (e.g., they fail to appear

[14] See here for a schedule of the bond courts in Cook County. Defendants are generally processed at CCJ after they have their bond hearing at the court.

[15] Active is defined as having at least 500 cases within a year, excluding days where a judge saw fewer than 40 cases. In the full data, this filter removes 3% of observations but 90% of unique judges. Between 2010 and 2016, two active judges are recorded working in bond court on the same date on less than 10 days out of over 2,500, about 0.3% of observations.

in court).[16]

IEM bonds (29%) allow defendants to leave jail at no cost (like an I-bond), but they are placed on electronic monitoring (EM). The defendant can be released off of EM if they pay 10% of the bond amount (like a D-bond). Prior to IEM bonds, EM could be coupled with a D-bond (D-EM) such that defendants were in jail until they paid 10% of the bond amount and were released onto EM. See Appendix A.1.1 for more details on bond types and background on IEM's introduction.[17]

Treatment Definitions I map these bond types and release statuses into three ordered pretrial treatments ($S \in \{1, 2, 3\}$). At the lowest level, defendants can be "released" ($S = 1$) if they fully exit the sheriff's office's custody through an I-bond, or because they were given an IEM or D-bond but paid the bond amount and were released within 7 days. The middle level is being on "EM" ($S = 2$), meaning the defendant was assigned an IEM bond but was not recorded as being released from it (e.g., paid the 10% amount) within 7 days. At the highest level, the defendant can be "detained" ($S = 3$) if they are given a D-bond and are not released from jail within 7 days. For the main results, the cutoff is 7 days, which is the 39th percentile for the duration in jail/EM, but 75% of releases happen within 3 days for D- and IEM bonds. This cutoff is largely arbitrary, and I construct robustness checks using alternative cutoffs. In particular, using the common 3-day cutoff yields similar results.[18]

Pretrial Period and Case Outcomes After the bond hearing, all defendants have the opportunity to plead guilty or proceed with the case (which can involve a future plea).[19] Prior to the case ending, through a trial, dismissal, plea, or being dropped, defendants can be rearrested or charged with new crimes and fail to appear in court.

[16]While defendants are liable for the bond amount, most defendants cannot pay the large sums, and cash bail has been shown to be ineffective at ensuring court appearances ([106]), and the threat of court fines has been shown to be ineffective ([107]) as the collection of such fines is rare ([108]).

[17]Judges also can but rarely do deny bond (1.4%) if they determine the defendant is a flight risk or a threat to the public. There are two additional bond types in the data A and C bonds, but they account for a minute share of bonds.

[18]Notably, 30% of defendants classified as 'detained' are eventually released before their case ends, though this number is comparable to that of other work, such as [30].

[19]The next steps depend on if the case is a misdemeanor or felony case. Felony cases proceed to a hearing which determines if the case can proceed with felony charges (usually a preliminary hearing, grand jury, or an information) and be transferred to the criminal division; otherwise, it is dropped or proceeds with misdemeanor charges. The full evolution of a case involves many events, and a flowchart for felony and misdemeanor cases can be seen in Figures A.1 and A.2, respectively.

Lengthy cases are common, and trials are rare, with 93% of cases with a guilty outcome involving a guilty plea. If guilty, the defendant can be sentenced to pay a fine, time served, or incarceration in prison (Illinois Department of Corrections) or in Cook County Jail. Following the case outcome and subsequent incarceration period, defendants can be rearrested or have a new case against them (a new case post-trial).

1.2.2 Background on Electronic Monitoring

EM Conditions and Details

While electronic monitoring can refer to a variety of technologies, all operate as electronic individualized surveillance systems with a similar mechanism and purpose. In this paper, I refer to the system operated by the Cook County Sheriff's Office between 2013 and 2015, which generally involved defendants wearing a radio-frequency ankle bracelet at all times. EM was used to ensure defendants did not leave their homes except at preapproved dates and times (e.g., a specific work or education schedule) or for a small set of "one-time movement" conditions.[20] This system is referred to as EM coupled with house arrest, and it is common across the United States ([4]). The ankle bracelet communicates with a monitoring unit installed in the defendant's home, which informs the sheriff' office of out-of-bounds movement and tampering.[21]

Defendants on EM agree to warrantless searches of their residence while on EM, and the possession of firearms, drugs, or contraband is a violation of EM bond conditions ([111]). As stated by the sheriff's office's EM information sheet, violations or noncompliance with the terms carry the "risk of criminal prosecution and re-incarceration," and damaged or missing equipment can be

[20] See this information sheet for the rules and information sheet for the EM program in 2020. While the domain of EM monitoring is generally one's home, exceptions can be made in advance for work, school, or other reasons ([8]), it but requires 2 days prior approval. The Sheriff's EM program is by far the most common form of EM ([109], [8]). See Appendix A.1.1 for a discussion of other EM programs. Time spent on EM in the Sheriff's program counts as days served in jail and thus can reduce one's time required to be served if found guilty ([8]). In many jurisdictions, EM can require defendants to pay a fee, but this was not common for pretrial EM during the period of study based on available sources. See [110] for images of the 2013-2015 system.

[21] While I do not observe whether or when the defendant has the EM system set up in their home, I do observe disposition codes in a defendant's case which provide more specific information on their EM status, such as explicitly stating a defendant was admitted into the Sheriff's EM program. I test the robustness of my results to modifying the treatment definitions using these disposition codes in Section 1.6.

charged as felony theft ([112]). Similar to other bail conditions of release (e.g., a condition of all bond releases is that the defendant appears in court), violations of the EM bond requirements can lead to re-incarceration in jail as well as additional charges.

EM as Middle Treatment

IEM bonds were introduced in June 2013 in the wake of a jail-overcrowding crisis and conflict between the court and the sheriff's office, and they quickly became popular among judges, comprising 28.7% of bonds in the following two and a half years, leading to a decline in other bond types. While IEM (EM) can be seen as a tool to reduce overcrowding in Cook County Jail, in practice, EM was used as a middle option between release and detention and applied to defendants who would otherwise have been released as well as those who would otherwise have been detained. Though no official guidance on EM as the 'middle' option existed at the time, the 2016 "Decision Making Framework" in Cook County explicitly places EM between release and detention (jail) ([113]).

Figure 1.2 shows that with the introduction of the IEM bond, mid-level D-bonds were replaced by IEM bonds. To demonstrate this more directly, I regress the probability of detention using defendant observables from 2009 to 2012; using these coefficients, I then construct an index for defendant severity (propensity to be detained) as the predicted probability of detention in the post-EM period for released, detained, and EM defendants. Figure 1.3 displays the distribution of the index for each treatment and shows that defendants assigned to EM are between those who were released and those who were detained based on their predicted likelihood of detention.

1.2.3 Data and Summary Statistics

Data

The main data for this study comes from the Circuit Court of Cook County and Cook County Jail. The court data contains information on defendants, cases, charge counts, and outcomes for almost all Cook County court cases between 1984 and 2019. This allows me to follow cases from inception to bond court, individual motions and case events, through to final case outcomes (e.g., a defendant demanding a trial, a guilty plea on a specific charge but not others, and a sentence to time served). I connect defendants across cases, allowing me to observe extensive criminal and case histories across tens of thousands of individuals over more than three decades, including past and future cases, arrests, guilty verdicts, charges, sentencing, and pretrial misconduct. I link the court data to data from the Cook County Jail, maintained by the Cook County Sheriff's Office. This jail data, spanning from 2000 to 2017, contains information on individuals' detention spells, intake, and release. I also link this data to Chicago Police Department arrest data.

While the collective data allows me to follow tens of thousands of defendants in Cook County through the criminal justice system from arrest in Chicago to sentencing over more than 15 years, this study focuses on two and a half years (July 2013 to December 2015) in which EM was one of the most common pretrial treatments for defendants. I focus on adult cases that went through Branch 1 with known bond types. The data is summarized at the booking and defendant level — which I will refer to as a "case" for simplicity.[22]

In order to quantify the intensive margin of defendants' recidivism (new charges after bond court) and avoid common issues with binary measures of recidivism ([114]), I quantify the social costs of different crimes by supplementing my main data with the crime cost information from [19].[23]

In the final data, I focus on completed cases within the Cook County court system that do not involve murder or felony sexual assault charges (in the current case) and have categorizable pretrial treatments. In the final data there are 84,332 defendant-bookings (cases) between July of 2013 and December of 2015 which passed through Branch 1, 51,348 of which comprise the main felony

[22]Linking these cases to jail spells leaves some cases with missing jail information. In the main analysis, I code missing releases for I-bonds as immediate releases, missing EM bonds as EM, and drop missing D-bonds. I test the robustness of my results to these codings in Section 1.6.

[23]I use Table 5 from [19], which contains the total tangible and quality of life costs associated with different crimes, to guide the construction crime costs for charges against defendants in the data as well as imputing costs for some unlisted crimes. A similar method is used in [115] to compute cost-weighted crimes per capita. However, because the cost of a single murder is so high (almost $8 million dollars, which is equivalent to about 200 police-reported robberies), I also conduct analyses using a "low" murder cost, making it equivalent to a police-reported rape (about $400,000) to determine if outlier defendants accused of murder in future cases are driving all of the results, similar to an exercise in [116].

sample.[24] See Appendix A.1.2 for more details on the data construction and filtering.

Summary Statistics

Table 1.1 displays summary statistics for defendant and case characteristics, and Table 1.2 displays summary statistics for a subset of outcomes by pretrial treatment, with release being the lowest treatment and detention being the highest.[25] As shown in Table 1.1, bond amounts, days in jail or on EM, and case durations are increasing in the treatment level. Black and male defendants (the majority of defendants) are more likely to receive a higher treatment level, and treatment level is increasing in the severity of charges against the defendant and in the severity of the defendant's case history increases. For example, released defendants are least likely to have a felony charge (45%) and have the fewest previous felony cases (1.5). In the middle, EM defendants are the most likely to have a felony charge (86%), but this is driven by drug felonies (e.g., 57% are charged with possession), and they have better case histories than detained defendants. Detained defendants are most likely to have violent felony charges (11%) and have the worst case histories.

For outcomes, released defendants are two and four times as likely to fail to appear (FTA) in court (14%) relative to EM and Detain defendants. The likelihood of a guilty felony charge increases with treatment level, from 22% for released defendants to 59% for detained defendants. Interestingly, EM defendants have the highest rates of having a new case pretrial (18%) relative to lower likelihoods for released and detained defendants (15% and 11%), likely due to detained defendants being released eventually (after 7 days) and EM defendants having new charges for bond violations. Post-trial, total new cases over four years are similar across treatments at around 1.6, though only about one-third of these are felony cases, and less than 5% are for violent felonies. The similarity across treatments is likely due to post-trial incapacitation in addition to selection and potential treatment effects.

[24] Though misdemeanor arrests are an important part of the criminal justice system ([117], [118]), EM was primarily used on felony cases. The main felony sample also excludes D-EM cases.

[25] Table A.1 and Table A.2 display the same for the felony sample. Table A.3 displays the means of all outcomes by treatment for the felony sample.

1.2.4 Framing Costs and Benefits

To understand to better frame the results, we can categorize the costs of each pretrial treatment (release, EM, detention) into four main categories: direct costs, the dollar cost of operating each treatment; crime costs, the social cost of crime under each treatment; punishment costs, the cost of enforcing each treatment; and indirect costs, all other spillovers onto the defendant or society. This paper provides estimates for changes in crime and punishment costs using new cases pretrial and indirect costs using new cases post-trial and sentencing outcomes, while direct cost estimates are taken from prior work. Furthermore, treatments differ in which mechanisms they rely on to reduce misconduct: the probability of detection (p) to deter misconduct or the level of incapacitation (d).

Direct Costs For direct costs, the comparisons are relatively straightforward and not studied directly in this paper. Jails cost about \$100-\$200 per detainee per day to maintain.[26] While pretrial EM is significantly cheaper than detention, monitoring defendants is still costly. The total cost of operation of EM is about \$15 per defendant-day in Cook County in 2021 ([113]), while release is almost costless and less expensive than EM. These estimates are average costs. Marginal costs, which may be more appropriate, are likely around 20% and 60% of average costs ([120]).[27] As the average defendant is in jail or on EM for over 100 days, this equates to marginal costs of jail between \$3,000 and \$9,000 for the average defendant.

Crime Costs Pretrial detention should reduce pretrial crime costs by incapacitating defendants in jail. On the other hand, EM operates through deterrence, with defendants' crimes or violations being more likely to be detected while on EM relative to release ($p^R < p^{EM}$), and through partial-incapacitation through house arrest conditions ($d^R < d^{EM} < d^D$). We can estimate changes in crime cost as the effect of EM versus release and detention on new cases pretrial; however, we must account for the fact that new cases pretrial depend on the probability of detection, which is higher for EM compared with release.

[26] For example, in 2011, CCJ cost about \$90 per defendant-day, while in 2021 it cost \$223 per defendant-day ([119]).

[27] Marginal costs are more difficult to calculate. [120] find that short-run and long-run marginal costs of incarceration are often around 20% and 60% of average costs. Thus, per defendant-day in jail, the average cost is around \$150, while the short-run and long-run marginal costs are around \$30 and \$90.

Punishment Costs Punishment costs result from defendants being punished, e.g., spending time in jail or being prosecuted or fined for pretrial misconduct. Naturally, detention involves high punishment costs as defendants effectively serve sentences while awaiting their case conclusion. And EM has lower punishment costs as house arrest is a less intense punishment than jail. However, because a defendant can be rearrested and placed in jail, release can lead to higher punishment costs if pretrial misconduct is detected and penalized (e.g., pretrial rearrests). By comparison, whether EM reduces punishment costs depends on whether being placed on EM increases or decreases the likelihood of receiving a new case pretrial. While EM aims to reduce pretrial crime through deterrence and partial incapacitation, it may result in higher punishment costs through two mechanisms.

First, because EM does not allow for perfect detection of defendants' misconduct, it applies sanctions to complementary but non-criminal activity (e.g., leaving one's home without permission), largely consistent with principal-agent models of multitasking ([56]). As such, while on EM, non-criminal activity can lead to re-incarceration, a new case, and subsequent punishments, inducing inefficiency and higher punishment costs. Any increase in new cases from violations can be seen as a cost of monitoring the agent (defendant) beyond the direct costs. If the rules of EM are too stringent, unforgiving, or unreasonable such that defendants are punished for non-criminal activity ([4]), then EM's punishment cost may be large.[28]

Second, if defendants are relatively inelastic (elasticity of crime, c, with respect to the probability of detection $-1 < \epsilon_p^c < 0$), and the incapacitation effect is relatively small, then defendants who are placed on EM will commit fewer crimes, but EM will still lead to increased pretrial rearrests. As a result, EM will increase punishment costs through heightened detection, partially offsetting gains through reduced crime.[29] Thus, the change in punishment cost depends on the effect of EM versus release on new cases pretrial.

[28]Notably, though most new cases pretrial under EM (77.76%) do not involve a violation charge, 55.17% of new cases with a violation charge include no other charges for low-level or serious crimes.

[29]Since the elasticity of x with respect to y is $\epsilon_y^x = \frac{y}{x}\frac{dx}{dy}$, and arrests (detected crime) $a = c \times p$, then $\epsilon_p^a = \frac{p}{a}\frac{da}{dp} = \frac{p}{c \times p}(c + p\frac{dc}{dp}) = 1 + \frac{p}{c}\frac{dc}{dp} = 1 + \epsilon_p^c$. So, $\epsilon_p^a < 0 \iff \epsilon_p^c < -1$. Furthermore, if defendants are elastic $\epsilon_p^c < -1$, then the punishment cost will decline under EM relative to release.

Indirect Costs We can measure some of the indirect costs associated with each treatment as the effect of EM versus release and detention on future recidivism (rearrests post-trial) and case outcomes (e.g., guilty findings). Beyond individual suffering, detention has high indirect costs by damaging employment, future income, and defendants' bargaining power leading to worse case outcomes.[30] Pretrial EM may be less coercive than detention and reduce long-run recidivism while causing less individual suffering. Relative to release, however, EM may be criminogenic (increasing future crime costs) or increase indirect costs by hurting defendants' bargaining power. Beyond this, EM may damage social ties, economic opportunities, and health outcomes and increase housing insecurity, as found in interviews with EM participants ([113]).[31]

1.3 Effect of EM vs. Release and Detention

1.3.1 Empirical Strategy

To identify the effect of EM versus release and EM versus detention, I instrument for the two endogenous treatments ($Release_{ic}$ and $Detain_{ic}$) by exploiting the quasi-random assignment of judges to a defendant-case (i, c). I interact bond court judges with defendant characteristics to instrument for pretrial treatment assignment because there are relatively few judges in the data (7) who see many cases (12,047 on average). This allows judges to have heterogeneous preferences over defendant observables ([38], [29], [31]). I construct the judge instruments by estimating:

$$1\{s_{ic} \in k\} = \beta_j^k 1\{Judge_{ic} = j\} X_{ic} + \kappa_{t(c)}^k + \psi_{ic}^k \tag{1.1}$$

where s_{ic} is the treatment to which defendant i in case c was assigned, $\kappa_{t(c)}^k$ are month and day of week (time) fixed effects for case c's bond court date, and β_j^k are judge-specific propensities for assigning defendants with characteristics X_{ic} (e.g., charges, past cases, race) to a pretrial treatment level within k. For example, $k = \{2, 3\}$ means the judge assigns the defendant to EM or

[30] See [29], [31], and [30]. Bargaining power is a significant factor in case outcomes ([121]). Spending time in jail may reduce future economic opportunity and increase criminal capital ([122], [123]).

[31] Alternatively, EM may reduce future crime through an individual deterrent effect (increasing defendants' expectations of punishment) — though the research on EM versus detention (which should have a similar effect) has not found evidence consistent with such a mechanism ([22]).

detention. Figure A.5 displays the heterogeneity in estimated β_j^k's across judges over defendant characteristics. I use fitted values from estimating equation (1.1) with OLS as the instrument for being assigned to a treatment within k, $Z_{ic}^k = \widehat{\Pr}\{s_{ic} \in k\}$.

Given the instruments, Z_{ic}^1 and Z_{ic}^3, I estimate the effect of EM versus release and EM versus detention with two-stage least squares (2SLS), with the following first and second stages:

$$[Release_{ic}, Detain_{ic}]' = \lambda_1 [Z_{ic}^1, Z_{ic}^3]' + \lambda_2 X_{ic} + \mu_{ic}. \tag{1.2}$$

$$Y_{ic} = \alpha_1 \widehat{Release}_{ic} + \alpha_2 \widehat{Detain}_{ic} + \theta X_{ic} + \epsilon_{ic} \tag{1.3}$$

where Y is the outcome of interest and $\widehat{Release}_{ic}$ and $\widehat{Detain}_{ic}$ are the instrumented treatments.

Instrument Validity

The validity of the instrumental variables strategy relies on multiple assumptions. First, we assume that the assigned judge is unrelated to the defendant, conditional on time fixed effects. The judge calendar (see Figure 1.1) alleviates concerns over violations of this assumption because judge assignment is determined solely by the sitting judge at the time of the bond hearing and not defendant characteristics. We can indirectly test this by seeing if defendant observables are predictive of judge assignment, which would suggest unobservables may be predictive as well, and thus exogeneity is violated. Table 1.3 displays the results of a multinomial logit with the defendant's bond court judge as the outcome and defendant observable as regressors (Column (2)) and month and day of week fixed effects as regressors (Column (3)). Consistent with exogeneity, defendant observables have virtually no predictive power (a pseudo-R^2 of almost 0). However, time fixed effects have significant predictive power (a pseudo-R^2 of almost 0.5), consistent with the calendar determining judge assignment.

An additional concern is that the judge influences defendant outcomes through means other than their pretrial treatment — for example, the judge also influences the defendant's assigned prosecutor — thus violating the exclusion restriction. Fortunately, Branch 1 operates on an entirely separate schedule from the other elements of the Cook County court system, and bond court judges do not play a role in future portions of the case — in the sample, only 0.2% of defendants saw their bond court judge in some capacity later in their case. Overall, these facts suggest violations of the exclusion restriction are not a significant concern.

Finally, we require that the judges actually influence treatment assignments. Table 1.4 displays the relationship between the standardized instruments, $\hat{Z}^1_{ic}$ and $\hat{Z}^3_{ic}$, and the endogenous variables $Release_{ic}$ and $Detain_{ic}$.[32] The first-stage relationship is strong: a 1 SD increase in $\hat{Z}^1_{ic}$ increases $Release_{ic}$ by about 9pp, while a 1 SD increase in $\hat{Z}^3_{ic}$ increases $Detain_{ic}$ by 10pp, and these relationships are consistent across specifications with and without controls and fixed effects. Figure 1.4 visualizes the support of $\hat{Z}^1_{ic}$ and $\hat{Z}^3_{ic}$ and the first stage.[33] Finally, Table A.4 contains results of the [124] robust weak instruments test, which indicate that weak instrument bias is not a significant concern.[34]

Interpreting 2SLS Results

2SLS does not generally recover the effect of one treatment versus another if there are multiple treatments without additional assumptions ([85], [87]).[35] We wish to know if $\hat{\alpha}_1$ and $\hat{\alpha}_2$ can be interpreted as the (weighted average of) causal effects of release versus EM and detention versus EM for some population of compliers — defendants whose treatment status was influenced by the

[32] Standardized instruments are first purged of the linear influence of defendant observables. For example, estimating equation (1.1) with $K = \{3\}$ and predicting the values gives Z^3_{ic}, then $\hat{Z}^3_{ic}$ can be recovered by taking the residuals from regressing $Z^3_{ic} = \beta X_{ic} + \kappa_{t(c)} + \gamma_{ic}$ ($\hat{\gamma}_{ic} = Z^3_{ic} - \hat{\beta} X_{ic} - \hat{\kappa}_{t(c)}$)and standardizing them ($\hat{Z}^3_{ic} = \frac{\hat{\gamma}_{ic} - \overline{\hat{\gamma}_{ic}}}{SD(\hat{\gamma}_{ic})}$). Similar results for the non-standardized instruments can be seen in Table A.6.

[33] Table A.5 displays the reduced form results. The judge instruments ($\hat{Z}^1$ and $\hat{Z}^3$) have a strong reduced form relationship with the outcomes of interest for pretrial misconduct; however, the strength of the reduced form relationship is either mixed or weak for case outcomes and post-case outcomes.

[34] Because instrumental variation comes from judges × observables = 154, we may have a many weak instruments problem. I use the specification $1\{s_{ic} \in k\} = \beta^k_j X_{ic} + \beta^k_2 X_{ic} + \kappa^k_{t(c)} + \psi^k_{ic}$ which produces one additional instrument per judge-characteristic interaction (which was previously collected into a single instrument). The [124] effective F-statistic for Z^1 is 35.8 and for Z^3 is 54.3, both are above the 2SLS critical values of 22.6 and 22.7, respectively, which reject the weak instruments null hypothesis with a worst-case bias greater than 5% of the OLS bias at a significance level of 5%.

[35] If treatment effects are constant across individuals, then 2SLS will return the average treatment effect even in the case of unordered treatments. However, rejection of the null in the over-identification test in Table A.4 is consistent with different judge instruments identifying different treatment effects, indicative of heterogeneous treatment effects. Furthermore, we cannot exploit the ordered structure to recover a meaningful average causal response of increased treatment ([125]) due to the treatments not being quantifiable in a meaningful sense.

judge instrument — or a local average treatment effect (LATE). [98] provide informal tests for the conditions under which 2SLS can recover a LATE. As discussed in Appendix A.1.3, I find small violations of these tests, so I cautiously proceed with the interpretation of the 2SLS results as a LATE.

1.3.2 Results: EM vs. Release

Column (1) of Table 1.5 displays the 2SLS results for the effect of a defendant being placed on EM relative to release for the main felony sample. Relative to release, EM reduces FTAs by -5.5pp ($p<0.05$) (-45.31% of release mean). This indicates that EM is an effective method for ensuring defendants show up for court, though this is consistent with a reminder effect rather than necessarily preventing flight, as [126] show that reminder text messages can have large effects on reducing FTAs and are not nearly as coercive as EM.

Despite the significant decrease in FTAs, EM *increases* the likelihood of a new case pretrial 7.1pp ($p<0.05$), which is a 41.5% increase relative to the release mean. We can decompose these effects into charge types: serious charges (violent crimes and non-drug felonies), low-level charges (drug crimes and non-felonies), and bond violation/escape charges (a single case can have multiple charge types). EM increases the likelihood of each charge type: 1.6pp ($p>0.1$) for serious, 3.5pp ($p<0.05$) for low-level, 4pp ($p<0.05$) for violations. As discussed in Section 1.2.4, the increase in arrests for crimes pretrial consistent with EM increasing the probability of detection, such that more crimes may be detected under EM while total crimes may actually decrease due to deterrence and incapacitation. Despite the increase in new cases (the extensive margin), on the intensive margin (measured by the dollar cost of new cases), EM has a noisy but negative effect. So, while EM increases arrests, this is largely driven by low-cost crimes. On the other hand, the increase in violation charges is consistent with EM as a "punitive surveillance" device with many inflexible rules by which defendants cannot reasonably abide ([127], [6], [4], [18]), leading to higher punishment costs.

I find that EM has a small or null effect on case outcomes, with a small and not statistically significant increase in the likelihood of a guilty felony charge and a decrease in the likelihood of being sentenced to incarceration. Similarly, EM has small positive effects but not statistically significant effects on total new cases over 4 years for all and felony cases, and on the intensive margin, EM has a noisy effect on post-trial cost crime over 3 years and over 4 years post-bond court. These results are consistent with EM being no more coercive or criminogenic than release, though they may be null due to the weakness reduced form relationship for these outcomes. Overall, EM does not appear to have higher indirect costs relative to release.

1.3.3 Results: EM vs. Detention

Column (2) in Table 1.5 displays the 2SLS results for the effect of a defendant being placed on EM relative to detention for the main felony sample. Relative to being detained pretrial, EM results in a significant increase in FTAs, 7pp ($p<0.05$) (241.7% increase relative to the detain mean). Similarly, EM also increases the likelihood of having a new case pretrial, by 16pp ($p<0.05$) (150.26% increase relative to the detain mean). Furthermore, there is a large and uniform increase across all charge types, meaning the increase in new cases pretrial is not solely driven by violations. EM decreases the intensive margin (cost of pretrial crime) when using the full murder cost (-\$8,300 ($p>0.1$)). However, this is due to outlier murder charges because when using the low-murder cost, EM leads to a statistically significant increase in pretrial crime costs (\$2,600 ($p<0.05$)). These results suggest that EM's deterrence effect and partial-incapacitation effect are not complete substitutes for detention's incapacitation effect and allow for defendants to FTA, but the increase in crime is relatively low-cost and likely less than even the short-run marginal cost of detaining a defendant.

EM has a small negative effect on the likelihood of being found guilty on a felony charge (-1.6pp ($p>0.1$)). However, EM does result in better case outcomes, with the likelihood of being sentenced to incarceration declining by -4.9pp ($p<0.05$), corresponding to -17.23% of the EM mean. These results suggest EM improves some case outcomes for defendants relative to detention and has lower indirect costs.

Despite a reduction in post-trial incarceration, EM has small and noisy effects on total cases and felony cases post-trial. EM has a negative but noisy effect on post-trial crime-cost when using both full (-\$17,000 ($p>0.1$)) and low (-\$1,800 ($p>0.1$)) murder cost, as well as crime-cost over 4 years following bond court for full (-\$40,000 ($p<0.05$) and low murder (-\$71 ($p>0.1$)). This suggests that any reduction in post-trial incapacitation or deterrence due to being placed on EM relative to detention is not sufficient to outweigh EM's lower criminogenic effect, and thus EM appears to be less criminogenic than detention resulting in either no change in or cost savings in post-trial criminal activity, further supporting EM's lower indirect costs relative to detention.

1.4 Heterogenous Effects

The 2SLS results do not tell us how the expansion of EM would affect different types of defendants, which is particularly important considering how the introduction or expansion of EM will lead to different populations being assigned to EM rather than release or detention. For this, we turn to a marginal treatment effects (MTE) framework. In this section, I build on [102]'s generalized ordered choice [128]-style model to identify MTEs under weaker assumptions. Then, I provide a tractable method for semiparametrically estimating said MTEs. With MTEs, we are able to construct relevant treatment parameters, such as the average effect of moving a released defendant to EM. This section focuses on the case of 3 ordered treatments while Appendix A.3 provides the general identification result and estimation method.

1.4.1 Empirical Strategy

Model

Selection into Treatment A bond court judge is randomly assigned (conditional on time fixed effects) to a defendant and assigns the defendant to one of three mutually exclusive treatment levels ($S \in \{1, 2, 3\}$) ordered by their intensity: release ($S = 1$), EM ($S = 2$), and detention ($S = 3$). The judge's decision is based partially on the defendant's observable characteristics (X). Different judges can view observable characteristics differently (e.g., one judge is harsh on drug offenders, another lenient, another prefers EM for them), so we interact the judge with observables to construct an instrument $Z = \text{Judge} \times X$. We can capture the judge's benefit from or desire to assign a defendant to higher treatment with some function $\tau(X, Z)$.

We assume all unobservable characteristics (unobservable to the econometrician but observed by the judge) that influence the judge's decision can be summarized in a single "latent" index V, which has a continuous but unspecified distribution. V captures the cost to the judge of assigning the defendant to a higher treatment and can be seen as the defendant's resistance to higher treatment, according to judges, or just resistance to treatment. We make no assumptions on the factors which determine V; V could be a function of the likelihood of committing a future crime, wealth, or hair length.

Collectively, the "net" benefit from assigning a defendant to a higher treatment level to the judge is $\tau(X, Z) - V$, and this index moves across a series of thresholds which are a function of judge preferences over observables ($C^s(Z)$). Higher treatments have weakly higher threshold values ($C^{s-1}(Z) \leq C^s(Z)\ \forall s$), and no one can be assigned below $S = 1$ nor above $S = 3$, so $C^0(Z) = -\infty$ and $C^3 = +\infty$. A defendant is assigned to treatment level s if and only if the net benefit is between the relevant cutoffs:

$$D_s = 1(S = s) = 1[C^{s-1}(Z) < \tau(X, Z) - V \leq C^s(Z)] \tag{1.4}$$

where $D_s \in \{0, 1\}$ indicates whether the defendant was assigned to treatment level s. This selection model assumes strict monotonicity of unobserved defendant's index (V), in that all judges agree on the ordering of V, and that higher values of V, all else equal, should receive weakly higher treatments.[36] However, judges can set different cutoffs for which values of V (conditional on X) receive higher treatments.[37]

[36] All judges agree that any defendant a with $V_a = v'$ should receive a *weakly* higher treatment than any defendant b who is identical to defendant a except that $V_b = v$ if $v' < v$. This is analogous to the monotonicity assumption discussed in [129] and [130], but in an ordered environment ([131]).

[37] For example, judge A is more lenient with felony drug charges than judge B on the EM (2) versus detention (3) margin, but more strict than judge B on the EM (2) versus release (1) margin: $C^2(A \times \text{Drug}) > C^2(B \times \text{Drug})$ and $C^1(A \times \text{Drug}) < C^1(B \times \text{Drug})$, so a larger range of values $\tau(X, Z) - V$ for defendants with drug charges will be assigned to EM under judge A relative to judge B.

Assuming V is independent of Z and X and using a standard transformation from [102], we can reorganize the selection model into a function predicted probabilities, and we can transform V, whose distribution we do not know, into a uniform random variable. Let $F_V(V) = U \sim Unif[0,1]$ ($F_V(\cdot)$ is the CDF of V), such that U now captures resistance to higher treatment. Let $\pi_s(Z,X)$ as the probability a defendant receives a treatment *higher than s* given Z and X. Whereas we did not know $\tau(X,Z)$ or $C^s(Z) \forall\ s$, we can estimate $\pi_s(Z,X)$ as predicted probability of higher treatment based on Z and X. $\pi_1(Z,X) = \Pr(S > 1)$ is the probability of not being released, and $\pi_2(Z,X) = \Pr(S > 2) = \Pr(S = 3)$ is the probability of being detained, given values of X and Z. Equation (1.4) can now be expressed as:[38]

$$S = \begin{cases} 1 \text{ (Release)}, & \text{if } U \geq \pi_1(Z,X) \\ 2 \text{ (EM)}, & \text{if } \pi_1(Z,X) > U \geq \pi_2(Z,X) \\ 3 \text{ (Detain)}, & \text{if } \pi_2(Z,X) > U \end{cases}$$

Essentially, the defendant is assigned to treatment higher than s if the observable factors pushing them higher, as measured $\pi_s(Z,X)$, are larger than their unobserved resistance to treatment, as measured by U.

Potential Outcomes Let Y be the outcome of interest for a specific defendant, such as whether the defendant fails to appear in court ($Y \in \{0,1\}$) or the number of cases against them post-trial ($Y \in \{0,1,2,\ldots\}$). Y_s is the defendant's treatment (s)-specific potential outcome: their outcome *if* they were placed in treatment level s ([132]). Then, the observed outcome of a defendant is their potential outcome for the treatment they actually received: $Y = \sum_{s=1}^{3} Y_s \times D_s$, where Y_s is a function of observables (X) and unobservable factors $\omega_s\ \forall s$.

For example, if we could observe both Y_1 and Y_2 for a defendant, then $Y_2 - Y_1$ would give us that defendant's treatment effect of being placed on EM ($S = 2$) relative to being released ($S = 1$). However, we only ever observe a single potential outcome for each defendant. Furthermore,

[38] Because $\pi_s(Z,X) = \Pr(S > s|Z,X) = F_V(\tau(X,Z) - C^s(Z)) = \Pr(\tau(X,Z) - V \geq C^s(Z))$, then $1(S = s) = 1[F_V(\tau(X,Z) - C^{s-1}(Z)) > F_V(V) \geq F_V(\tau(X,Z) - C^s(Z))]$, so $1(S = s) = 1[\pi_{s-1}(Z,X) > U \geq \pi_s(Z,X)]$.

$\mathbb{E}[Y_s] \neq \mathbb{E}[Y|D_s = 1]$ because treatment is not randomly assigned and is possibly assigned based on potential outcomes.

For estimation and identification in this section, I assume full independence of observables and unobservables ([133]), $(X, Z) \perp (\omega_1, \omega_2, \omega_3, U)$, and that Y_s is linear in observables. This implies additive separability: $Y_s = \beta_s X + \omega_s$, meaning the way observables influence potential outcomes is unrelated to the way unobservables do.[39]

Treatment Effects

Because all judges agree on the ranking of defendants in terms of U but disagree on where to set cutoffs and how to weigh observable factors, two nearly identical defendants in terms of U can be assigned to different treatments solely due to their quasi-randomly assigned judge. The judge only influences outcomes by influencing treatment, not by influencing potential outcomes. Given we observe X, which influences both treatment and potential outcomes, we must overcome the fact that the unobserved component, ω_s, may also influence treatment assignment in order to measure the effect of one treatment versus another. As we have assumed V (and thus U) captures all unobserved factors determining treatment, we have that: $\mathbb{E}[Y_s|X = x, U = u] = \beta_s x + \mathbb{E}[\omega_s|u]$.

So, we will use the variation in judges to study how defendants across the distribution of U respond to different treatments. Because U ranks defendants by how likely they are to be assigned to a higher or lower treatment, different treatment effects across different values of U inform how outcomes will change if, for example, EM is expanded to defendants who are likely to be detained (low U, all else equal) or released (high U, all else equal). Note that we can only make such comparisons where there is common support, meaning that judges must have sufficiently different cutoffs such that we observe individuals with $U = u$ in different treatments.

MTEs and MTRs We can define the marginal treatment effect (MTE) as the average treatment effect of a defendant with observables $X = x$ and unobservable resistance $U = u$ from one

[39]While additive separability can be obtained with weaker assumptions than full independence ([95]), assuming full independence is appropriate in this case because of the interaction between observables and judges in constructing the instruments.

treatment s to another s' as: $MTE_{s',s}(X = x, U = u) = \mathbb{E}[Y_{s'} - Y_s|X = x, U = u]$.[40] In particular, we are interested in the effects of moving from release to EM ($MTE_{2,1}(x,u)$) and moving from detention to EM ($MTE_{2,3}(x,u)$). We can decompose any MTE into the difference between two marginal treatment response (MTR) functions, $\mathbb{E}[Y_{s'}|X,U]$ and $\mathbb{E}[Y_s|X,U]$ ([103]). Given the prior assumptions, we have:

$$
\begin{aligned}
MTE_{s',s}(X,U) &= \mathbb{E}[Y_{s'} - Y_s|X,U] = \underbrace{\mathbb{E}[Y_{s'}|X,U]}_{\text{MTR of } s'} - \underbrace{\mathbb{E}[Y_s|X,U]}_{\text{MTR of } s} \\
&= (\beta_{s'} - \beta_s)X + \mathbb{E}[\omega_{s'}|U] - \mathbb{E}[\omega_s|U]
\end{aligned}
\tag{1.5}
$$

Then, we can identify MTEs by first identifying MTRs separately. In Appendix A.3, I show that each MTR can be identified relying solely on variation between π_s and π_{s-1} for $\mathbb{E}[Y_s|X,U]$ $\forall s$. For example, by taking the derivative of $\mathbb{E}[Y \times D_s|\pi_s, \pi_{s-1}, X]$ (which can be estimated) with respect to π_{s-1}, we recover the MTR for $S = s$, $\mathbb{E}[Y_s|U = u, X = x]$, at $u = \pi_{s-1}$. [101] developed a similar approach but with the aim to bound MTRs.

Treatment Parameters of Interest As shown in [91], with marginal treatment effects and full support, we can construct a variety of treatment parameters of interest: in particular, the average treatment effects of moving from release or detention to EM ($ATE_{2,1}$ and $ATE_{2,3}$), the average treatment effect of moving from release to EM for released defendants (ATR), and the average of moving from detention to EM for detained defendants (ATD).

The ATE tells us the expected difference in an outcome if we took a random defendant in the population and assigned them to one treatment (EM) versus another (release or detention), fixing x at $\bar{X}$ (the mean of observables across all observations). Because u is uniformly distributed in the population, each point on the MTE is weighted equally.

$$
\begin{aligned}
ATE_{2,1} &= \int_0^1 MTE_{2,1}(u, \bar{X})du, \\
ATE_{2,3} &= \int_0^1 MTE_{2,3}(u, \bar{X})du
\end{aligned}
$$

[40] When the treatment levels are adjacent (e.g., $s' = s + 1$), the MTE is referred to as the transition-specific marginal treatment effects (TSMTE) ([134].

The ATR tells us the expected difference in outcomes if we took a random defendant who was released and assigned them to EM — that is, what would happen if we removed release as an option for defendants and assigned them to EM instead? Similarly, the ATD tells us the expected difference in outcomes if we took a random defendant who was detained and assigned them to EM — that is, what would happen if we removed detention as an option for defendants and assigned them to EM instead? These effects are relevant when discussing policies under which EM is expanded to replace release or detention.[41]

$$ATR = \int_0^1 MTE_{EM,R}(u,x)W_R(u,x)du,$$
$$ATD = \int_0^1 MTE_{EM,D}(u,x)W_D(u,x)du$$

In contrast with the ATE, which evenly weights all points on the MTE curve because u is uniform in the population, the ATR(D) applies heterogeneous weights. The ATR(D) weighs the effect of the observables of individuals more likely to be released (detained) more heavily, meaning those with low (high) values of π_1 (π_2), and they weigh the effects of unobservables more heavily for individuals who are unobservably more likely to be released (detained), equating to higher weights on high (low) values of $U = \pi_1$ ($U = \pi_2$). These weights are both contained in W_R and W_D, which both integrate to 1, though their exact composition depends on functional form assumptions in practice.

Estimation

Estimation of MTRs requires first recovering π_1 and π_2 and then recovering MTEs as the difference between MTRs. Steps and assumptions for estimation are discussed in more detail in Appendix A.3.

Recovering π π_1 and π_2 can be recovered separately by estimating equation (1.1) with the dependent variables being $1[S_i > 1] = D_{i,2} + D_{i,3}$ and $1[S_i > 2] = D_{i,3}$, respectively, using a

[41] However, they differ from the policy relevant treatment effect focusing on shifting propensity scores and treatment-uptake decisions ([91], [133], [97]).

probit model and predicting treatment probabilities.[42]

Recovering MTRs and MTEs As discussed above, we take the derivative of $\mathbb{E}[Y \times D_s|\pi_s, \pi_{s-1}, X]$ with respect to either π_s or π_{s-1} to recover an MTR ($\mathbb{E}[Y_s|X, U]$) at $\pi_s, \pi_{s-1} = u$. Given additive separability of Y_s, $\mathbb{E}[Y_s|X = x, U = u] = \beta_s x + \mathbb{E}[\omega_s|U = u]$.

Using the fact that we observe Y and $D_s = 1[\pi_s \leq U < \pi_{s-1}]$, we can construct $Y \times D_s$: $Y_i \times D_{i,s}$ is 0 if $S_i \neq s$ and Y_i if $S_i = s$. I define $\Lambda_s(\pi) = \int_0^{\pi} \mathbb{E}[\omega_s|U = u]du$, such that:

$$
\begin{aligned}
\mathbb{E}[Y \times D_s|\pi_s, \pi_{s-1}, X] &= \beta_s X \times (\pi_{s-1} - \pi_s) + \mathbb{E}[\omega_s \times D_s|\pi_s, \pi_{s-1}] \\
&= \beta_s X \times (\pi_{s-1} - \pi_s) + \Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s)
\end{aligned}
$$

where the change in the second line is due to $\int_{\pi_s}^{\pi_{s-1}} \mathbb{E}[\omega_s|U = u]du = \int_0^{\pi_{s-1}} \mathbb{E}[\omega_s|U = u]du - \int_0^{\pi_s} \mathbb{E}[\omega_s|U = u]du$. Note that the separable nature of $\Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s)$ is a result of the single index model.

We can approximate the difference $\Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s)$ using sieves which can be, for example, a B-spline or polynomial. Specifically, let $\Phi_s(\pi)$ denote a vector of basis functions used to approximate $\Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s)$, such that $\Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s) \approx \phi_s[\Phi_s(\pi_{s-1}) - \Phi_s(\pi_s)]$, where ϕ_s is a vector of coefficients for each s. For example, with a 3rd polynomial sieve, $\Phi_s(\pi) = [\pi, \pi^2, \pi^3]$ for each s. Using this sieve approximation and including ϵ as an error term, we can estimate with OLS:

$$
Y \times D_s = \beta_s X(\pi_{s-1} - \pi_s) + \phi_s[\Phi_s(\pi_{s-1}) - \Phi_s(\pi_s)] + \epsilon \tag{1.6}
$$

Then, we can approximate the MTR by taking the derivative with respect to π_s or π_{s-1} to recover $\mathbb{E}[Y_s|x, u] \approx \hat{\beta}_s x + \hat{\phi}_s \Phi'_s(u)$. Finally, we can take the difference between MTRs evaluated at the same values of X and U to recover $MTE_{s+1,s}(x, u) \approx (\hat{\beta}_{s+1} - \hat{\beta}_s)x + \hat{\phi}_{s+1}\Phi'_{s+1}(u) - \hat{\phi}_s \Phi'_s(u)$.

[42]This allows for all judges' preferences over observables to be heterogeneous across treatment margins (approximating instrument-dependent $C^1(Z)$ and $C^2(Z)$), but also allows observables to uniformly affect the likelihood of crossing a treatment threshold ($\tau(X, Z)$). Note that this is similar to, but not equivalent to, estimating an ordered probit with judge-specific thresholds, as it involves additivity across judge × characteristics by threshold and is less sensitive to misspecification of the distribution of V.

Main results use a 3rd degree polynomial for Φ, meaning MTRs and MTEs effectively use a 2nd degree polynomial. 95% confidence intervals are based on bootstrapped estimates with 200 runs producing non-symmetric confidence bands (see Appendix A.3.5 for more information).

Common Support MTEs can be identified where common support exists between MTRs. Because there are only 3 treatment levels, common support hinges upon the only middle treatment, EM, where we require overlapping support for π_1 and π_2 in the EM sample.[43] Figures A.7 - A.9 display the support and variation of π_1 and π_2. Using a 1% sample trim, the support of EM is $\pi_1, \pi_2 \in [0.25, 0.77]$ (see Figure A.9), so the main results focus on EM versus release and EM versus detention $u \in [0.25, 0.77]$, corresponding to mid-range defendants and excluding defendants at both extremes.

Constructing Treatment Parameters I follow [99] in the construction of these treatment effects (see Appendix A.1.4 for details). Importantly, the ATR, ATD, and ATEs can only be identified with full support (the support of integration is 0 to 1). Given that we lack full common support ($u \in [0.25, 0.77]$), we instead can identify common support equivalents: CATEs, CATR, CATD (i.e., common support ATEs, ATR, ATD) ([133], [41]).[44] In practice, I rescale weights to ensure they integrate to 1 over the common support. See Figure A.10 for the weights for the unobservable components under common and full support.

Comparison With Prior Methods

This semiparametric sieve approach has advantages over prior methods for estimating MTEs, specifically the fully parametric approach that assumes jointly normal unobservables (the "Normal" method) from [102], [134], and [104]. In addition to being fast and straightforward, this approach allows for non-linear MTEs, while the Normal method produces MTEs with monotonic and largely pre-determined shapes (effectively linear for all but the more extreme values of u) as

[43] Because $\pi_0 = 1$ and $\pi_3 = 0$, MTRs for release ($E[Y_1|x, u]$) and detention ($E[Y_3|x, u]$) only require variation in π_1 and π_2, respectively.

[44] Note that with [41]'s notation, we would refer to these analogous parameters simply the ATE/ATR/ATD for the common support sample, while [133] use $\widetilde{ATE}$, for example, to differentiate the common support ATE from the true ATE.

a result of the parametric assumptions. While prior applications of the Normal method only recover MTEs, this method also allows us to recover MTRs. Furthermore, while an advantage of the Normal method is that it allows for MTE estimation across the full support, the sieve method can be easily adapted to a fully parametric polynomial model to recover MTRs beyond common support and make full support comparisons. In Section 1.6, I compare the main common support results to full support using polynomial and the Normal method. Appendix A.1.5 provides more comparisons between the method presented in this paper and the Normal method for MTEs.

1.4.2 Results: EM vs. Release

Columns (1) and (2) of Table 1.6 and the left panels of Figures 1.5 - 1.7 display the effects of EM relative to release for the average defendant (CATE) and released defendants (CATR), along with 95% confidence intervals. Figures A.11 - A.13 display the MTE results, where higher values of U (x-axis) correspond to defendants who are easier to release.[45]

Relative to release, EM decreases the likelihood of a failure to appear in court. The effect is largest for low U defendants and smallest for high U defendants, and this selection pattern is consistent with judges assigning EM to defendants who would benefit (in terms of lower FTAs) from EM relative to release, based on unobservable factors. As a result, the CATE estimate (-9.8pp (p<0.01)) is larger than the CATR (-5.8pp (p<0.01)) (close to the 2SLS estimate), meaning expanding EM to defendants more likely to be released results in smaller benefits in terms of FTA reduction.

EM increases new cases pretrial for less resistant defendants (less likely to be released), but the effect moves negative as resistance increases, and this results in a noisy and near null effect across the distribution despite a positive and large 2SLS estimate. The noisy effect is driven by a null effect on new case charges for serious crimes, a negative effect on low-level charges, and a positive effect on violation charges– all sharing the same selection pattern with larger effects

[45]Figures A.17-A.19 display the MTRs. Recall that the MTEs are evaluated at values of $U = \pi_1$, so the x-axis is a measure of unobserved resistance to higher treatment — for example, $MTE_{2,1}$ at $U = \pi_1 = 0.25$ is for individuals who are *unobservably* harder to release (less resistant) and at $U = \pi_1 = 0.77$ are unobservably easier to release (more resistant).

for low-resistance defendants and smaller or negative effects on high-resistance ones. Overall, EM reduces the likelihood of charges for low-level crimes with -3pp ($p<0.1$) and increases the likelihood of violations by 3.2pp ($p<0.01$) for the average defendant. The effects on defendants likely to be released (CATR) are more muted, with a -1.8pp ($p>0.1$) decrease for low-level charges and a 1.4pp ($p<0.05$) increase for violations. Consistent with this, the effects on pretrial case cost are negative but noisy. From this, we can conclude that EM likely has a crime-reducing effect relative to release, reducing the likelihood of a new case with crime-related charges, but there is no significant effect on the intensity or social cost associated with these charges.

For case outcomes, EM increases the likelihood of a guilty felony charge and being incarcerated for defendants least likely to be released, but this effect decreases rapidly as defendants become more likely to be released. This results in positive CATEs (6.1pp ($p<0.05$) and 15pp ($p<0.01$)) but a negative CATR for felony charges (-2.8pp ($p>0.1$)) and a barely positive effect on sentencing (1.8pp ($p>0.1$)). While these patterns could be indicative of EM being coercive for higher-severity defendants and improving outcomes for low-severity defendants — possibly by reducing FTAs and rearrests and thus strengthening their cases — these results are largely suggestive due to small 2SLS estimates and the weak reduced form relationship.

Post-trial, EM increases the total number of new cases over 4 years, with a parabolic selection pattern: small effects on high and low resistance defendants and larger effects for mid-resistance defendants. This results in similar CATE and CATR estimates (0.39 ($p<0.01$) and 0.35 ($p<0.01$)) slightly larger than the 2SLS estimate. However, the effects on felony cases are small (and null for 2SLS), and EM has a noisy but near null (or positive in the case of low-murder cost) effect on total case cost over 4 years post-bond court, suggesting no significant criminogenic effect.

Overall, relative to release, EM increases punishment costs through increasing new cases for violations, and there is weak evidence for higher indirect costs through increased future low-cost recidivism and worse sentencing. However, EM reduces crime costs through a reduction in new pretrial cases for criminal activity and lower pretrial crime costs, as well as a reduction in FTAs. However, this reduction in pretrial new cases is small and driven by low-level crime. Selection patterns indicate that both the costs and benefits of moving from release to EM are smaller for defendants who are released relative to the average defendant.

1.4.3 Results: EM vs. Detention

Columns (3) and (4) of Table 1.6 and the right panels of Figures 1.5 - 1.7 display the CATE and CATD estimates along with 95% confidence intervals, and Figures A.14 - A.16 display the MTE results.[46]

Relative to detention, EM increases the likelihood a defendant fails to appear in court across all defendants, but the effect decreases in magnitude as defendant resistance increases. So, defendants who are unobservably easier to detain experience smaller increases in FTAs, a pattern inconsistent with selection into treatment upon gains in terms of reduced FTAs. This results in a lower CATD (3.3pp ($p<0.01$)) than CATE (5.3pp ($p<0.01$)), both smaller than the 2SLS estimate.

For new cases pretrial, EM increases the likelihood of a new case for the average defendant (6.5pp ($p<0.01$)) and similarly for defendants more likely to be detained (7.3pp ($p<0.01$)). This effect is driven by low-level and violation charges with CATEs (CATDs) of 3.8pp ($p<0.05$) (3.1pp ($p<0.1$)) and 3.7pp ($p<0.01$) (5.3pp ($p<0.01$)), respectively. Importantly, defendants assigned to EM relative to detention are no more likely to have new cases with serious charges. Notably, EM's effect on violations is relatively constant across defendant resistance. These results are in contrast with the 2SLS estimates, which are uniformly positive and larger than the CATEs or CATDs, and suggest over-weighting low-resistance defendants if interpreted as a LATE. Nevertheless, the effect on pretrial crime costs is noisy and negative using full murder cost but small and positive with low-murder costs — consistent with the small effects on serious charges. Overall, it appears EM does not result in more costly pretrial crime but does involve more rearrests for violations.

The MTE results suggest that EM improves case outcomes relative to detention, decreasing the likelihood of a guilty felony charge and of being sentenced to incarceration, with effects increasing

[46] Figures A.17-A.19 display the MTRs. Recall that the MTEs are evaluated at values of $U = \pi_2$, so the x-axis is a measure of unobserved resistance to higher treatment — for example, $MTE_{2,3}$ at $U = \pi_2 = 0.25$ is for individuals who are *unobservably* easier to detain (less resistant) and at $U = \pi_2 = 0.77$ are unobservably harder to detain (more resistant).

in magnitude as defendant resistance to detention increases. This is consistent with judges having some signal over defendant guilt or case strength when determining pretrial treatment, such that defendants who are unobservably (to the econometrician) less resistant to detention experience smaller changes in outcomes. These patterns result in CATEs of -30pp ($p<0.01$) and -29pp ($p<0.01$) and CATRs of -22pp ($p<0.01$) and -23pp ($p<0.01$), respectively; however, the 2SLS estimates are significantly smaller (though still negative) than the CATD and CATE estimates.[47]

Post-trial, EM has a positive effect on total new cases over 4 years for low-resistance defendants and has a negative effect for high-resistance defendants, with a similar pattern for felony cases. While this results in null CATE estimates (similar to the 2SLS estimates), the CATDs are positive (0.28 ($p<0.01$) for all and 0.14 ($p<0.01$) for felony). However, when post-trial cases are weighted by their crime cost, EM has a noisy negative effect on the average (-\$44,000 ($p<0.1$)) and detained (-\$31,000 ($p>0.1$)) defendant with full murder cost, and small effects with low murder cost. On net, EM has a negative effect on total case cost over 4 years following bond court for both full and low murder cost, with effects slightly increasing in magnitude as defendant resistance increases. This results in full and low murder cost CATEs (CATDs) of -\$63,000 ($p<0.01$) (-\$52,000 ($p<0.05$)) and -\$5,700 ($p<0.05$) (-\$1,400 ($p>0.1$)), with similar 2SLS estimates.

Overall, EM allows for increased pretrial misconduct in both pretrial crime, driven by low-level charges and violations, and FTAs. However, the main estimates are consistent with EM improving defendant case outcomes (though estimates range widely) and having a small effect on future cost-weighted recidivism. Selection patterns indicate that the average defendant experiences larger benefits and smaller costs of being placed on EM relative to detention compared with defendants who are detained.

[47] While these are large estimates, they are not inconsistent with prior work: [30]'s 2SLS estimates for the effect of release versus detention show release reduces guilty verdicts by about 24% of the detained defendant mean.

1.5 Policy Implications

1.5.1 Are Defendants Elastic?

As discussed in Section 1.2.4, EM aims to reduce pretrial crime through deterrence (increasing the probability of detection) and partial incapacitation (house arrest). If EM solely operated through incapacitation, and the change in detection rates was small, then EM's effect on new cases pretrial would be proportional to EM's effect on pretrial crime. However, the effect on new cases pretrial, which was relatively small relative to EM versus release, will understate the effect on pretrial crime. This is because EM likely increases the probability of detection, and thus a larger share of crimes will be detected under EM. If EM solely operated through incapacitation (house arrest), then this would suggest EM's crime-reduction effect is small. We can explore the sensitivity of pretrial crime with respect to the probability of detection and EM's crime-reduction effect.

First, we can infer a lower bound for the average defendant's elasticity with respect to the probability of detection, $\epsilon_p^c \approx \frac{\frac{\Delta c}{\frac{c^{EM}+c^R}{2}}}{\frac{\Delta p}{\frac{p^{EM}+p^R}{2}}}$, where $\Delta x = x^{EM} - x^R$, using the fact that $arrests\ (a) = crime\ (c) \times p$, to give us an expression for implied elasticity.[48] Even though we do not observe p^R or p^{EM}, we can trace the implied elasticities by assuming reasonable values for p^R across values of p^{EM}, and we observe a^R and a^{EM} as the average MTRs (weighted for average or released defendants) for a specific crime type, and similarly for fitted values using 2SLS estimates.[49]

The results are displayed in Figure 1.8 for low-level crimes and serious crimes using low, middle, and high values for p^R. The results indicate that low-level crime is likely elastic with respect to surveillance $\epsilon_p^{c,low} < -1$, though defendants more likely to be released are slightly

[48] This is a lower bound on $\epsilon_p^c \geq \hat{\epsilon}_p^c$ because any incapacitation effect (reducing c^{EM}) will be attributed to deterrence.

$$\epsilon_p^c \approx \hat{\epsilon}_p^c = \frac{\frac{\Delta c}{\frac{c'+c}{2}}}{\frac{\Delta p}{\frac{p'+p}{2}}} = \frac{\frac{\frac{a'}{p'}-\frac{a}{p}}{\frac{a'}{p'}+\frac{a}{p}}}{\frac{p'+p}{p'-p}} = \frac{\frac{p'+p}{p'-p}\left(\frac{a'}{p'}-\frac{a}{p}\right)}{\frac{a'}{p'}+\frac{a}{p}} = \frac{\frac{2p+\Delta p}{\Delta p}\left(\frac{\Delta a+a}{\Delta p+p}-\frac{a}{p}\right)}{\frac{\Delta a+a}{\Delta p+p}+\frac{a}{p}}$$

[49] Based on FBI statistics, over 40% of violent crimes are cleared, with about 30% of robberies cleared, and almost 20% of property crimes are cleared. I assume slightly higher than average clearance rates for crimes committed by defendants who are released on bail because their location is known and they were in recent contact with the system and are tracked by the court system and the Sheriff (e.g., court attendance). See `https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/clearances` for 2017 clearance rates.

less elastic. However, 2SLS results suggest defendants are likely relatively inelastic, consistent with the positive point estimates for new cases pretrial. We cannot reject the null that serious crime is relatively inelastic with respect to the probability of detection ($\epsilon_p^{c,serious} > -1$) across all estimate types (2SLS, CATE, CATR). The potential inelasticity of crime with respect to detection reduces the efficacy of surveillance as a social savings policy because even though crime is reduced, socially costly arrests may rise, thereby increasing punishment costs. Furthermore, these are both the lower bounds for elasticities, meaning defendants are likely more *inelastic* than $\hat{\epsilon}_p^c$ suggests.

The result that defendants may be relatively inelastic in serious crimes is in contrast with other empirical studies in which increased detection led to lower *observed* violations ([135], [136], [49]) with which the effects on low-level crimes are more consistent. However, the inelasticity for serious crimes is consistent with the discussions of arrest rates and incarceration ([64], [65]).[50] This raises an important policy issue with respect to surveillance: if surveillance detects more crimes but punishments stay high, and defendants are not sufficiently elastic with respect to detection, then surveillance will result in less crime but more punishment.[51]

1.5.2 How Much Crime Does Surveillance Prevent?

Second, we can use a similar method to compute the total cost of crime prevented under EM relative to release for the average defendant. Using a similar logic as above, and using $c \times g$ (cost of crime) $= \frac{a \times g}{p}$, so $\Delta cg = \frac{(a \times g)^{EM}}{p^{EM}} - \frac{(a \times g)^{R}}{p^{R}}$. This form simplifies cost-weighted crime as a single type with a single probability of detection, rather than summing over different types with different probabilities of detection. With this simplification, we can use average MTRs ($(a \times g)^R$ and $(a \times g)^{EM}$) for the crime cost of new cases (arrests) pretrial, and similarly for CATR weights.[52] Note that this calculation does not require any assumption on EM's incapacitation effect.

[50] Even if longer prison sentences deter crime, if potential criminals are relatively inelastic, then the incarcerated population will rise despite higher deterrence.

[51] Relatedly, in the theoretical literature discussing racial disparities in hit rates, it has been argued that the optimal search rate is not that which produces the highest hit rates but that which deters the most crime (i.e., officers should target the more elastic group) ([66], [67]).

[52] 2SLS estimates predict negative crime cost under EM, making the exercise more difficult to perform.

The top row of Figure 1.9 displays the implied changes in pretrial crime due to EM's crime-reduction effect for all charges. Assuming a crime-cost weighted (mainly weighted by murder) $p^{EM} = 0.5$, the point estimates suggest EM prevents nearly \$25,000 of pretrial crime for the average defendant and less than \$10,000 for the released defendants — importantly, we cannot reject a null (or even positive) effect by EM relative to release. However, this assumes that EM detects 100% of crimes and that all charges are correct. Given that a large fraction of charges does not result in a guilty finding, we can re-estimate the curve using guilty charges only, and the effects shrink significantly: the average defendant effect is less than \$10,000, and the released defendant effect is less than \$7,500 for $p^R = 0.5$ and assuming EM detects all crimes. Only the guilty CATR estimates can reject a null effect but cannot reject small gains. To avoid results being driven by outlier-murder cases, we can use the low-murder cost, which significantly reduces the gains to less than \$4,000 for average and \$3,000 for released with $p^R = 0.5$ and $p^{EM} = 1$. For comparison, the direct cost alone of EM operation for the average main sample defendant on EM (about 100 days) is about \$1,500.

1.5.3 Surveillance vs. Incapacitation

The results in the prior section suggest that EM is preferable to detention (better case outcomes, less costly total new cases both pre- and post-trial), except with respect to pretrial misconduct. The central question for EM versus detention is whether or not EM's crime-reducing effect and lower indirect and direct costs are a sufficient replacement for detention's incapacitation effect and higher direct and indirect costs. [30] estimate that pretrial detention (relative to release) costs over \$50,000 in lifetime costs per marginal defendant. While we cannot provide a similar calculation for EM versus detention due to a lack of individual financial data, we can compare the costs of pretrial misconduct under EM versus detention using a similar method as above.

Figure 1.10 displays the changes in pretrial crime cost between EM and detention (release following detention). Essentially, for EM's effect to be, on average, less effective over the pretrial period than detention, EM would need to detect *fewer* crimes than release (post-detention). This may be due to individuals released after long detention durations committing more serious crimes than those on EM due to jail's criminogenic effects.

By construction of the treatments, detained defendants may be released at some point during the pretrial period (just not within 7 days). So, we can examine the magnitude of how much crime EM allows for assuming detention is complete incapacitation, that is, defendants are held in jail until their case is complete, essentially being denied bond.[53] Figure 1.11 displays the results. For guilty charges, assuming $p^{EM} > 0.5$, crime costs on EM are less than \$5,000 for the average and detained defendant. For comparison, [19] estimates the cost of a larceny/theft at around $\$5,000$, while the cost of a single day of pretrial detention in Cook County is over \$100 more expensive than a day on EM– making the direct cost savings alone at least \$4,000 for the median defendant. Overall, it is highly unlikely that EM's effect is not a cost-effective substitute for detention's incapacitation effect.

1.5.4 Net Costs and Savings from Expanding EM

By applying dollar amounts to the various outcomes in the framework from Section 1.2.4, we can approximate the net savings associated with replacing release or detention for EM. Table 1.7 displays the breakdown of cost differences for ATEs, ATR, and ATD estimates using approximations for various costs.[54] Figure A.20 displays the total cost MTRs for both full and low-murder costs — notably, the MTRs are generally decreasing in defendant resistance to higher treatment.

In terms of direct cost differences replacing release with EM costs a few hundred dollars, while replacing EM with detention saves more than \$7,000 per defendant. Changes in failure to appear

[53] While useful, if implemented, detention would likely result in additionally harmful effects for defendant cases and recidivism outcomes (as well as employment loss), in addition to unmeasured costs on families, individuals, and communities, and an increase in large direct costs from holding more people in jails.

[54] I use average costs of \$15 per defendant-day for EM and \$150 defendant-day for jail and apply a 40% marginal cost. FTAs are valued at \$1,000. Crime costs are based on aggregated arrest costs assuming a 30% cost-weighted detection rate under release, post-detention release, or post-trial, and a 60% cost-weighted detection rate under EM. Pretrial crime costs include punishment costs. Sentencing costs are assumed to be \$20,000 per defendant sentenced to incarceration, based on [137] estimates of average prison cost per defendant in Illinois in 2015 of \$33,507 and applying a 60% marginal cost discount. This also assumes defendants sentenced to incarceration are serving one year on average, which may be an underestimate (over 75% of individuals sentenced to prison or jail are sentenced to prison).

rates are neither sizable nor costly enough to make a significant difference on average. Replacing EM with release leads to increased sentencing costs as defendants are more likely to be sentenced to incarceration, at $350 and $3,000 for average and released, while EM reduces sentencing costs relative to detention at $6,000 and $5,000 for average and detained.

While the prior estimates are relative precise, crime cost estimates are not. Differences in pre-trial crime cost are large with full murder costs with EM saving between $19,000 and $47,000 per defendant relative to both release and detention using full murder costs for all defendant types, but confidence intervals include zero and cannot reject modest costs or even more massive savings. Using low-murder costs, EM still saves between $8,000 and $12,000 per defendant relative to either release or detention, and all the confidence intervals reject savings lower than $2,000 per defendant. Recidivism (post-trial) crime costs over 3 years add to these savings for EM relative to detention with massive but imprecise savings with full murder cost (cannot reject losses) and smaller but precise (reject savings smaller than $4,500). However, recidivism costs reverse some of the benefits of EM relative to release. EM results in over $20,000 in post-trial crime with full murder costs (but cannot reject massive savings or costs). With low-murder costs EM costs about $10,000 per defendant in post-trial crime, with smaller confidence intervals, though only released defendants reject any savings.

On net, EM than release is more costly for released defendants ($6,000 and $2,500 using full and low-murder costs), but confidence intervals are wide. In contrast, the signs for average defendants disagree, and the confidence intervals include both massive savings and costs. Relative to detention, EM likely induces net savings as estimates are consistently negative for average and detained defendants, and confidence intervals using low-murder cost are much more precise and reject savings smaller than $4,500 per defendant. Notably, these estimates are biased in favor of higher treatment levels, as they do not account for employment, health, or personal costs, which are likely high in detention ([30]) if not EM as well. Nevertheless, these results are consistent with EM being a cost-effective alternative to detention, even for otherwise-detained defendants. At the same time, it is not clear that EM induces cost-savings relative to release, particularly for released defendants.

1.6 Robustness

I test the sensitivity of my results to alternative samples and specifications for five main outcomes. Figures 1.12-1.14 display the CATE, CATR, CATD, and 2SLS estimates, while Figures A.21-A.23 display the MTE curves. Additional checks are discussed in Appendix A.1.6.

Changing the Release Cutoff As discussed in the main results, the cutoff used for determining the treatment is 7 days. In order to test whether the results are driven by this arbitrary cutoff, I redefine the cutoff as 14 days and 3 days (as in [30] and [31]), and results are largely unaffected.

Felony Cases From Start In the main estimation, I include both felony and non-felony Branch 1 cases in constructing the judge instruments as that ensures all the cases the judge sees contribute to the instrument, while the results are for the felony sample only. I re-compute the instruments using only felony cases, and these results are generally consistent with the main results.

Including D-EM Bonds and Excluding Missing EM D-EM bonds (where the defendant pays the bond amount to be released from jail to EM) comprise a small portion of observations that were dropped in the felony sample. To see if this is changing the results, I re-add them to the sample and re-compute MTEs. Also, as discussed in the data section, EM bonds with missing durations are coded as EM. To see if that is influencing the results, I drop all missing duration EM bond observations. In both cases, the results do not deviate significantly from the main results.

Judge Instrument Only The main instruments rely on variation in judge $\times$ observables, but the random assignment is between cases and judges, conditional on time fixed effects. Relying instead only on judge fixed effects ("$Z(fe)$"), which results in a weaker instrument with smaller common support, produces results that are generally consistent with the main specification except with larger effect sizes — though occasionally the estimates are highly different (e.g., for FTAs and likelihood of incarceration for EM versus detention).

Full Support Parameters The treatment parameters were constructed over the region of common support, but we can compare these estimates to those using full support (e.g., an ATE rather than a CATE) using a sieve-style polynomial parametric assumption and the fully parametric 'Normal' method. These full support ATE, ATD, and ATRs generally agree with the main results, with disagreements for outcomes with weak reduced form relationships or when the Normal method extrapolates monotonically due to its parametric assumptions (e.g., FTAs for EM versus release).

Using Disposition Codes As mentioned in Section 1.2.1, defendants who receive EM bonds may not actually be admitted into the EM system. To determine if this potential miscoding of treatments will influence the results, I reclassify defendants' EM status based on disposition codes observed in their case history. I construct 4 alternative codings which allow for additional conditions under which the defendant is classified as on EM based on disposition codes. The results are displayed in Figure 1.13, and they are broadly consistent with the main results.

Re-coding Missing Treatments As discussed in Section 1.2, jail data is unmatched to cases for a subset of the sample. In order to test the sensitivity of my results to how these cases with missing jail data were coded or dropped, I redo the analysis with 4 additional samples in which all cases with missing jail data are kept, usual filters applied, and the entire analysis is redone (e.g., π's are re-computed for each sub-sample).[55] The 4 cases are combinations of coding all missing (jail data) EM bonds as EM or release and coding all missing D-bonds as detention or release. Overall, as shown in Figure 1.14, these re-codings do not alter the main interpretation of the main findings.

1.7 Conclusion

This paper explores the effect of pretrial electronic monitoring, an individualized surveillance technology, on defendant outcomes relative to both release and detention across defendant types. In this effort, I develop a general method for semiparametrically estimating marginal treatment effects in ordered treatment environments. Compared with detention, I find that EM allows for more low-level pretrial misconduct, but EM does reduce costly pretrial crime and leads to weakly better case outcomes and fewer high-cost crimes in the future. Compared with release, I find that EM prevents low-level pretrial crime and reduces failures to appear in court, but I cannot reject

[55]To reduce miscoding, I remove all D-EM cases prior to re-estimating the π's as well.

a small cost-weighted crime reduction effect. Overall, the evidence suggests that EM is a viable alternative to detention for higher-level defendants, but it is not clear that its potential benefits outweigh its costs relative to release for lower-level defendants.

Crucially, I find that EM causes an increase in new cases pretrial for bond violations compared with both release and detention, which highlights an additional cost of using imperfect monitoring to enforce compliance. Based on testimonials and research in Cook County and other municipalities using EM, the rules and requirements of EM are often too stringent and inflexible, potentially resulting in worse economic, social, and health outcomes ([113]). Compliance with the threat of significant punishment involves little leniency for mistakes, legitimate emergencies, big life events (e.g., moving apartments), or daily tasks (e.g., doing laundry, picking kids up at school) ([109], [6], [4], [138]). One improvement could be the relaxation of punishments for violations of EM and increased discretion and leniency, which would be consistent with not only the [71] model of optimal deterrence (where increasing detection can be met with a decrease in punishment) but also with successful "swift-and-certain" sanction regimes ([49], [50]).[56]

Finally, this paper calls for an increased understanding of the costs and benefits of modern surveillance technology within the criminal justice system and more broadly within economic spheres. The existing work on surveillance technology in economics focuses largely on mass surveillance and uses the lens of innovation, political economy, and social control ([80], [81], [82], [83]). In contrast, the results of this study are consistent with the criticisms of EM and similar *individualized* surveillance technologies in use across the United States ([15], [18]): defendants cannot successfully adjust to the level of monitoring, and this leads to increased criminalization and worse outcomes. Yet, this work is only a first step in understanding whether surveillance technologies are desirable in the criminal justice system.

[56]Furthermore, the results emphasize that the large costs are driven solely by outlier individuals charged with murder. This indicates that targeting resources and monitoring toward individuals who are at risk for committing murder could reduce serious crime while also reducing the scope of coercive policy on most defendants.

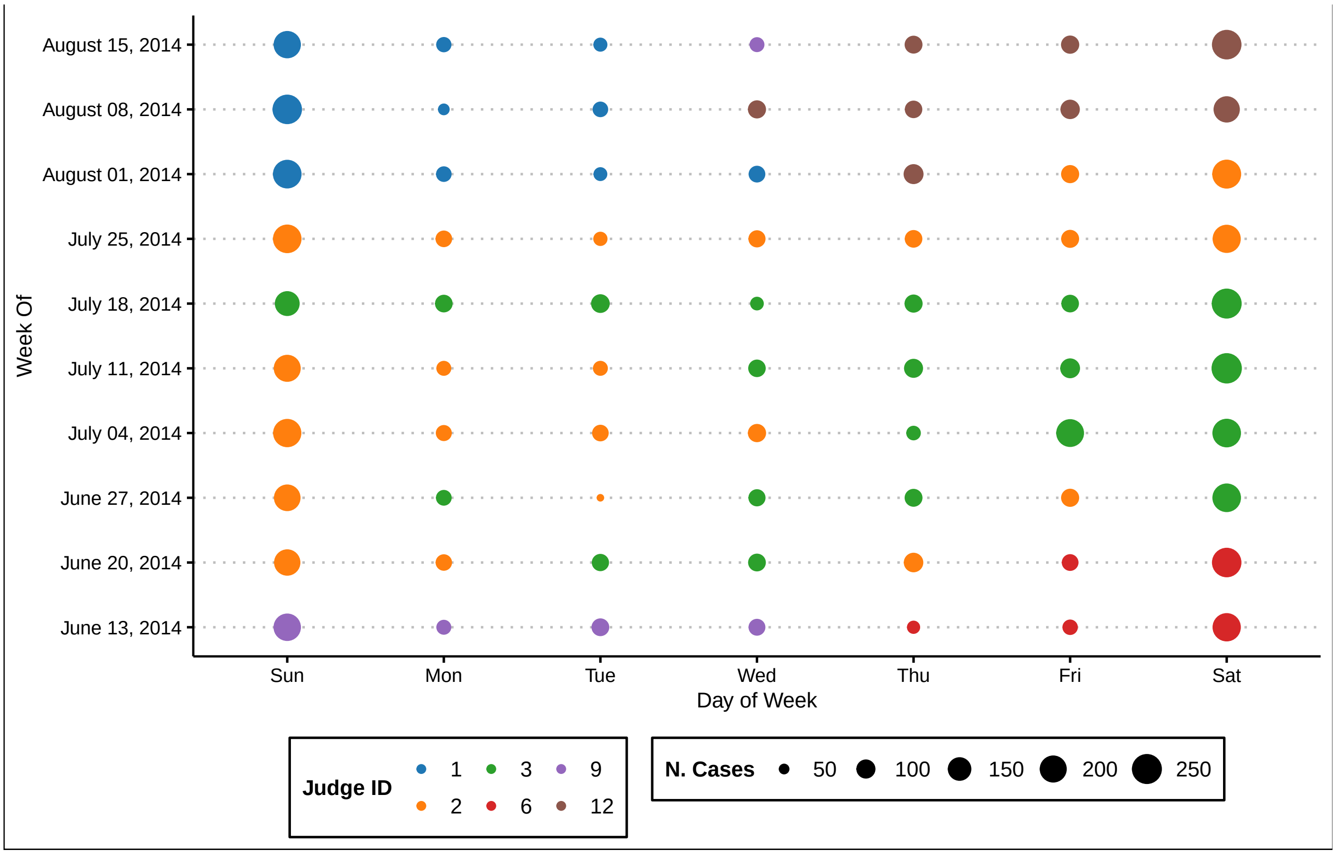


Figure 1.1: Example of Bond Court Judge Rotation Calendar

Note: Figure displays the caseloads of active judges in the Cook County Bond Court (Branch 1, Room 100) by week and day of the week between the weeks of June 13, 2014, and August 15, 2014. There were 6 active judges in this period (dot color), while the size of each dot denotes the number of cases they saw that day.

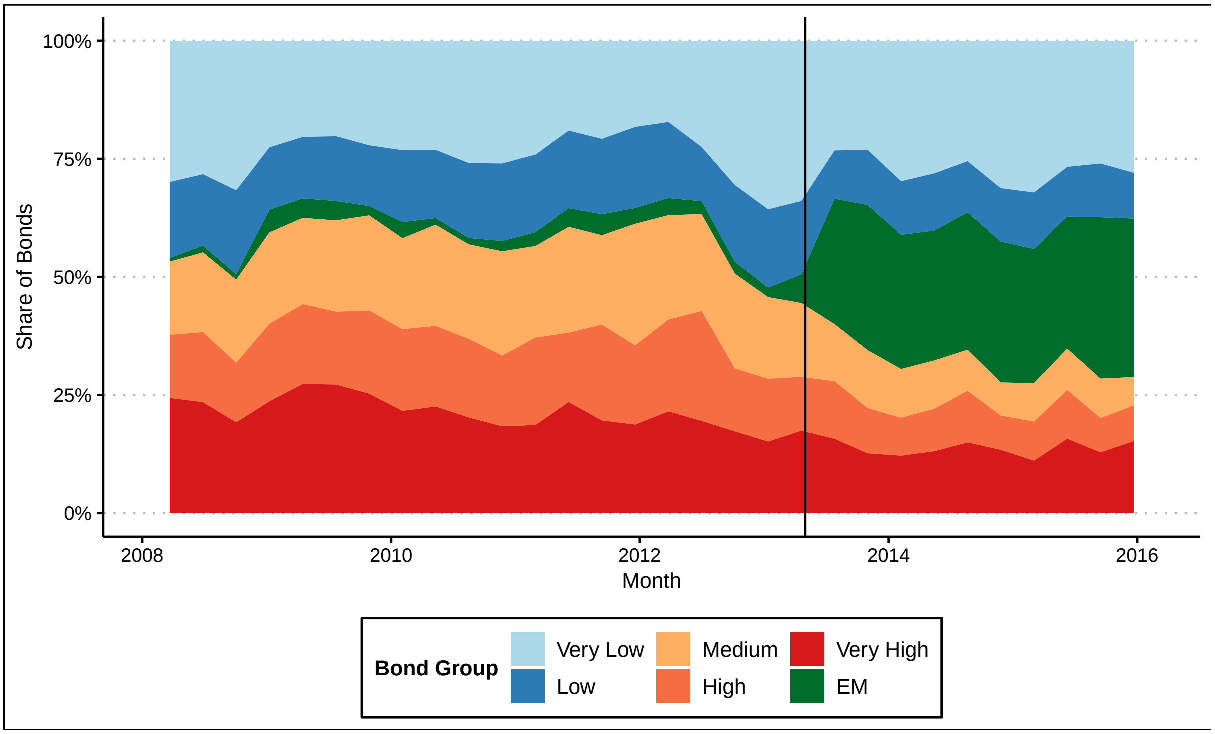


Figure 1.2: Bond Types Over Time

Note: Figure displays the composition of bond types within the sample between March 2008 and 2015 aggregated by the year-month of bond date. Bond group EM denotes IEM bonds, and other bond groups are determined by the bond price required for release (i.e., the bond price for any D-bond, \$0 for I-bonds, and +\$∞ for bond denial). Very low contains bonds with amounts between \$0 and \$7,500; low contains bonds with amounts between \$7,500 to \$20,000; medium contains bond with amounts between \$20,000 to \$40,000; high contains bonds with amounts between \$40,000 to \$60,000; and very high contains all bonds with amounts above \$60,000.

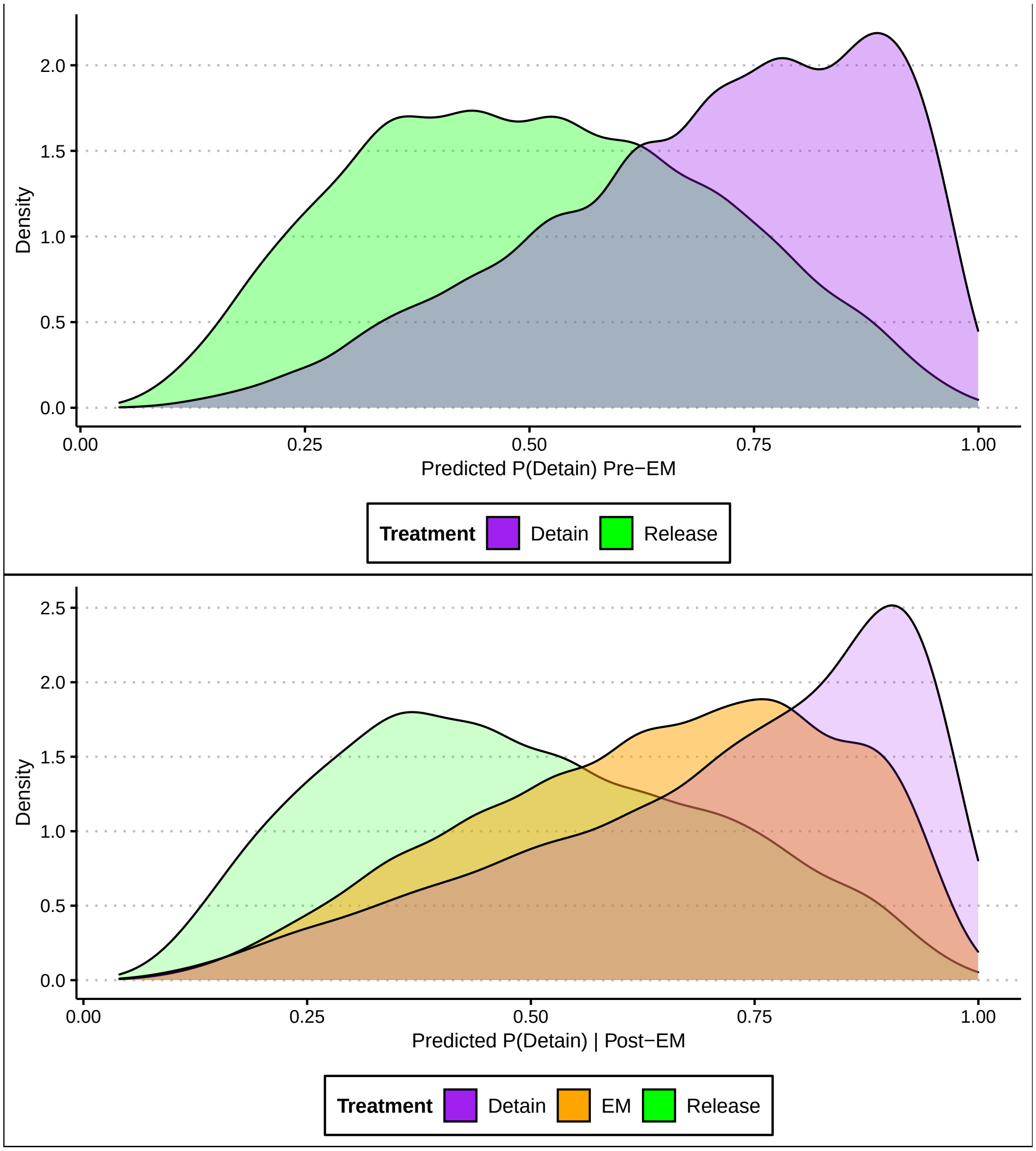


Figure 1.3: Distribution of Defendants by Treatment, Pre- and Post-IEM

Note: Figures display the distributions of pretrial treatments during the period (Detain, Release) for pre-IEM (2009-2012) and post-IEM (July 2013 - 2015). X-axis is the defendant's predicted likelihood of being detained in the pre-IEM period based on their case observables. Coefficients for predicting likelihood of detention are recovered from regressing detention on defendant observables in the Pre-IEM period, then predicted values are computed using the coefficients on data from the Pre-IEM period (top) and Post-IEM period (bottom).

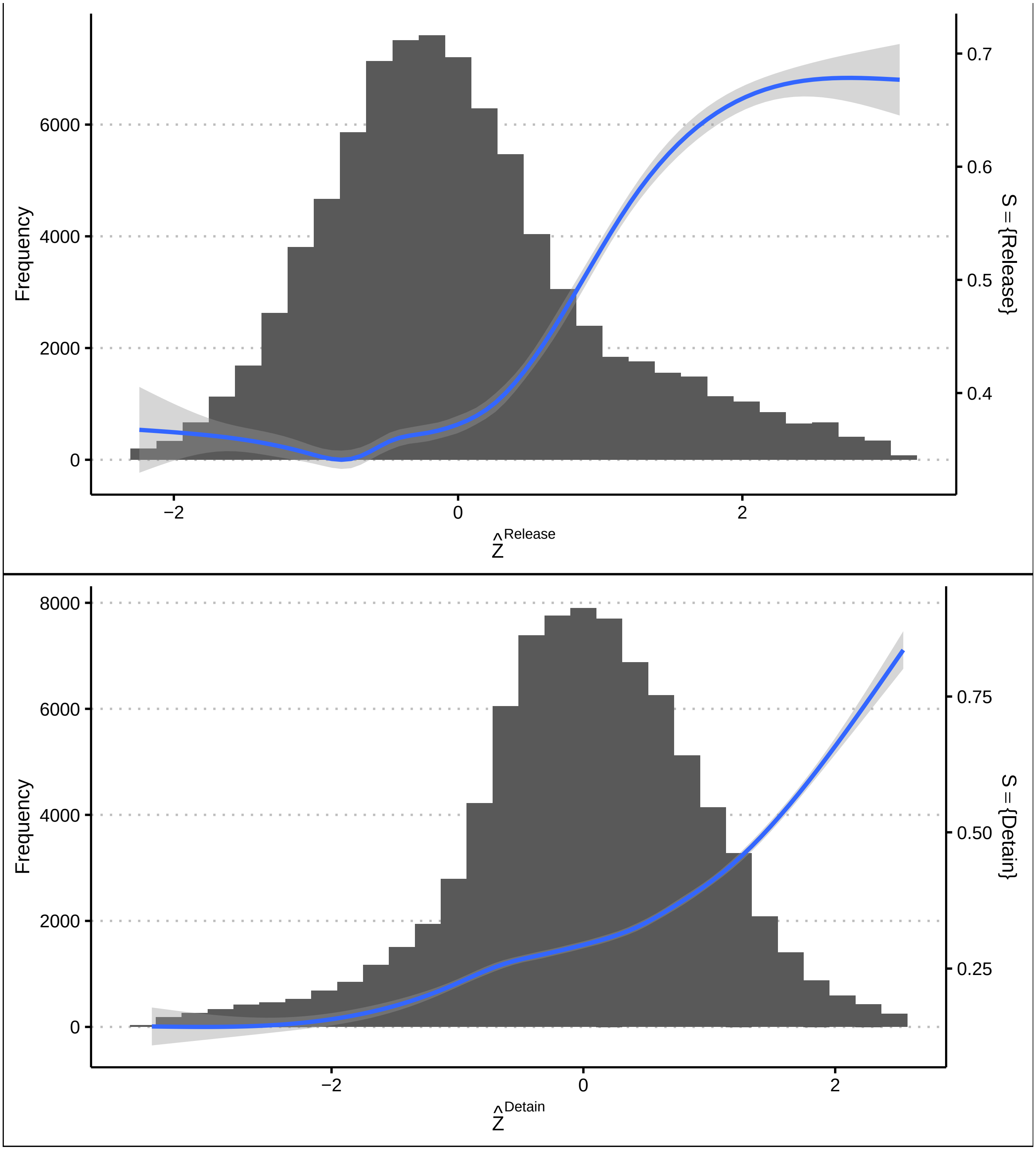


Figure 1.4: Distribution of Instrumented Treatment and First Stage

Note: Figure displays the local linear fit and support of the instrumented probabilities of a defendant being assigned to Release (top) and Detention (bottom). Instrumented probabilities are constructed using an LPM and are residualized to remove the linear influence of observables and time fixed effects, and are standardized. The x-axis is the value of the instrument (standardized); the left y-axis is the frequency of the instrument value; the right y-axis is the likelihood of the defendant being placed on the treatment of interest.

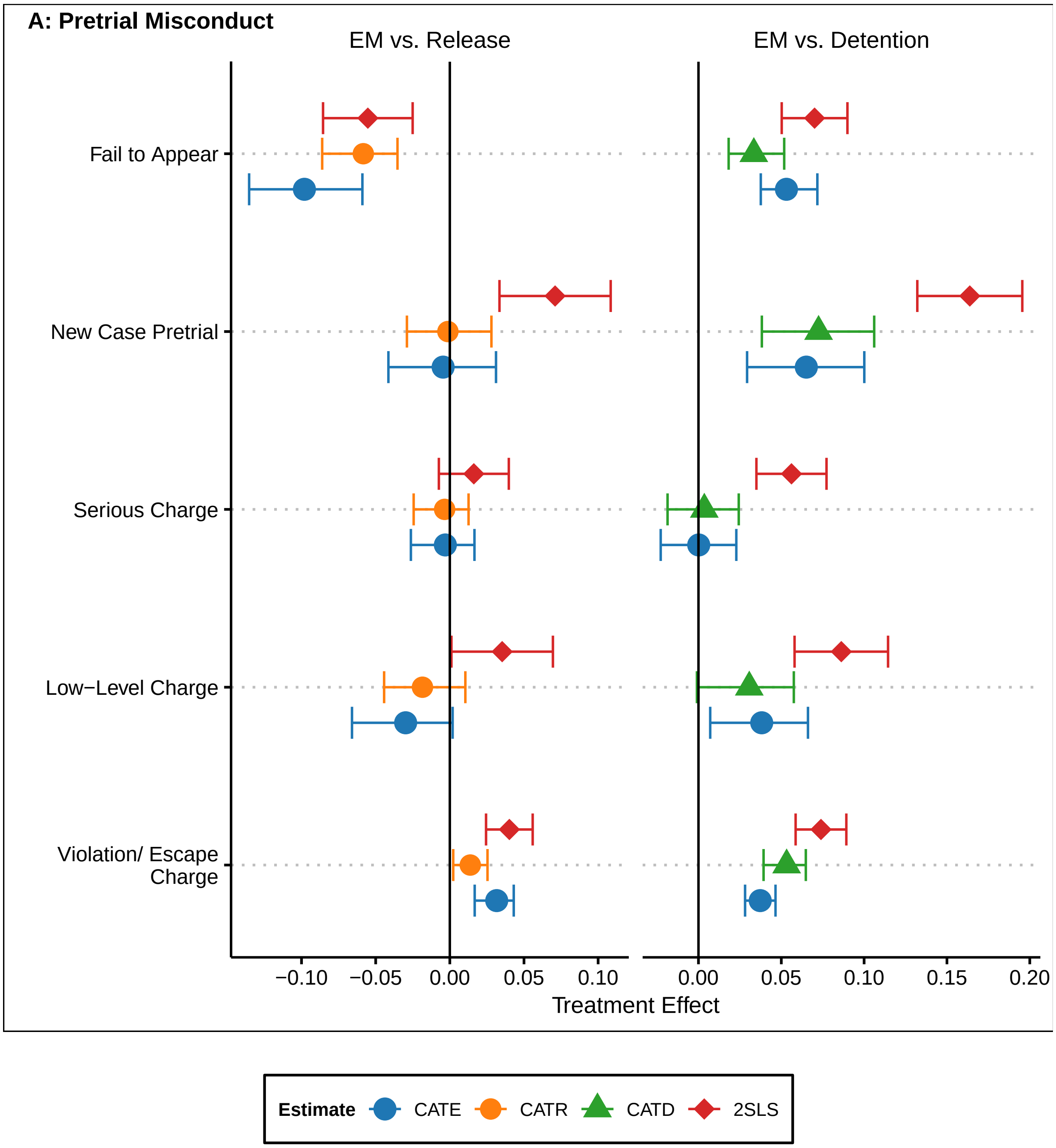


Figure 1.5: Treatment Effect Plots for Pretrial Misconduct

Note: Figure displays treatment effects for EM versus Release and EM versus Detention for main felony sample cases. Outcomes consist of binary variables relating to pretrial outcomes. ATE = average treatment effect, ATR = average treatment effect on the released, ATD = average treatment effect on the detained. (C) denotes parameter constructed using common support with the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs for CATE, CATR, and CATD, and standard errors clustered at the defendant level are used for the 2SLS estimates.

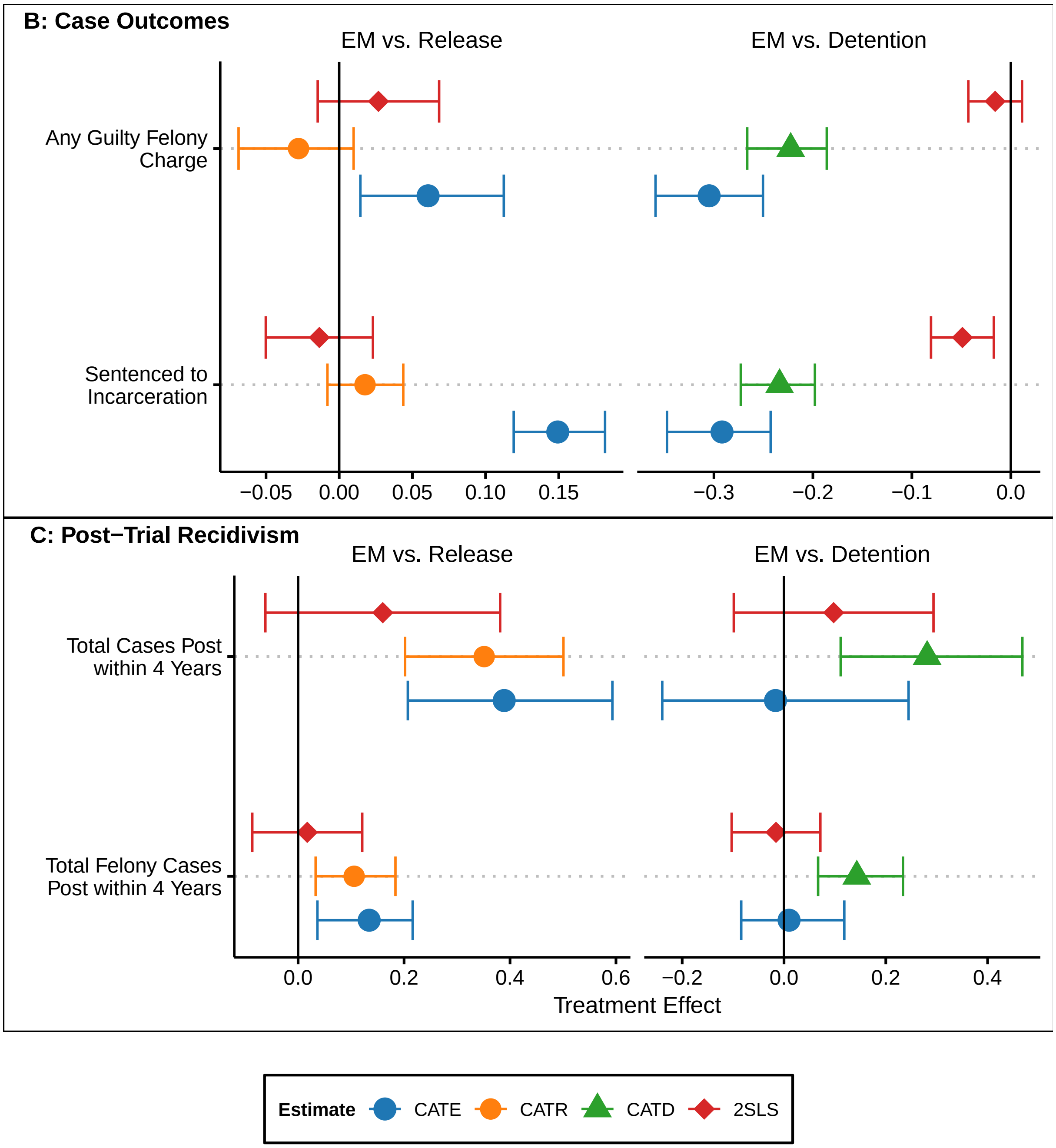


Figure 1.6: Treatment Effect Plots for Case and Post-Trial Outcomes

Note: Figure displays treatment effects for EM versus Release and EM versus Detention for main felony sample cases. Outcomes consist of variables relating to case and post-trial outcomes. ATE = average treatment effect, ATR = average treatment effect on the released, ATD = average treatment effect on the detained. (C) denotes parameter constructed using common support with the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs for CATE, CATR, and CATD, and standard errors clustered at the defendant level are used for the 2SLS estimates.

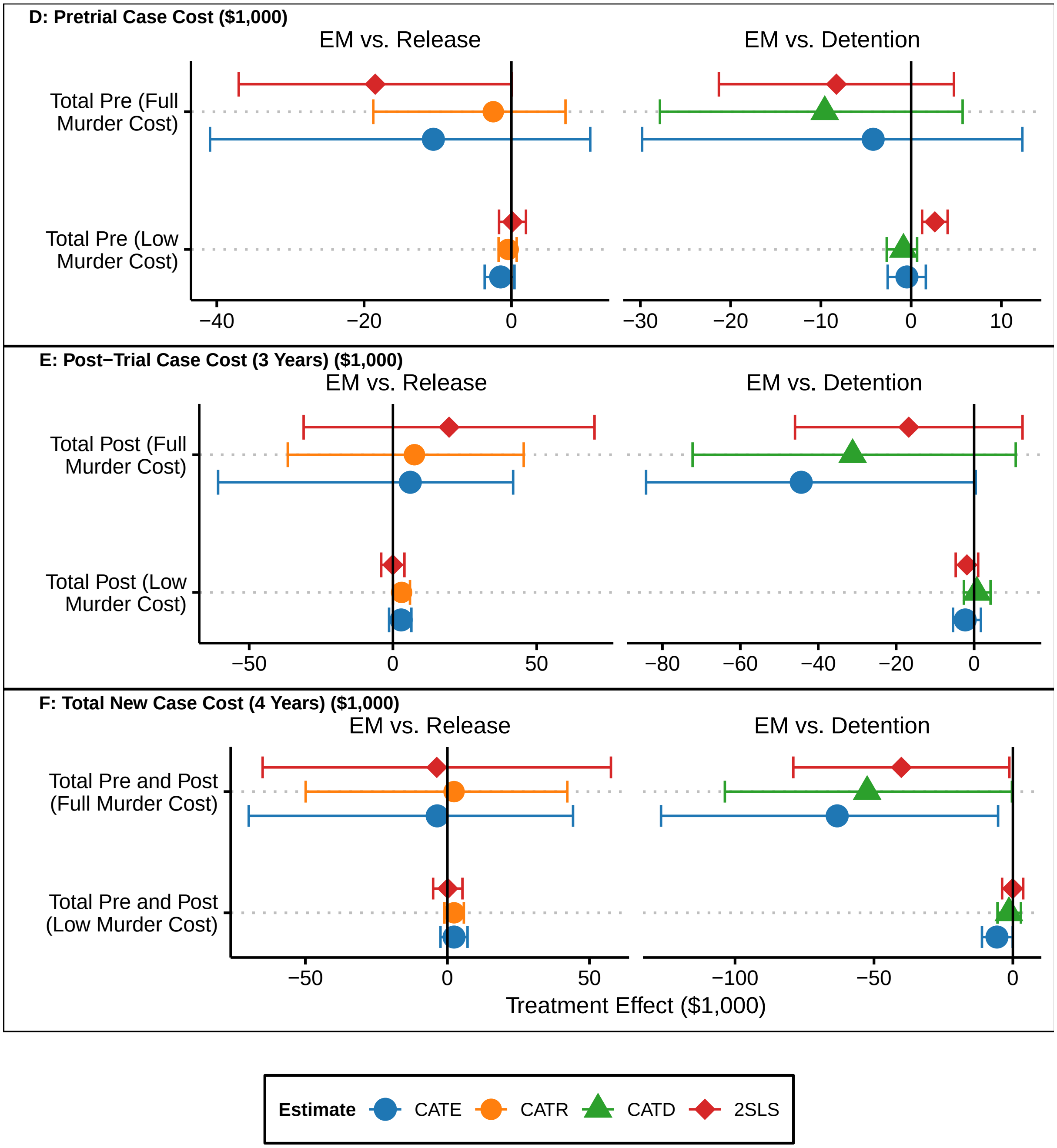


Figure 1.7: Treatment Effect Plots for New Case Costs

Note: Figure displays treatment effects for EM versus Release and EM versus Detention for main felony sample cases. Outcomes consist of variables pre- and post-trial new cases weighted by their incidence costs cost based on estimates from [19] (in which the full murder cost is around $8,000,000). Low murder cost uses a value of around $400,000. ATE = average treatment effect, ATR = average treatment effect on the released, ATD = average treatment effect on the detained. (C) denotes parameter constructed using common support with the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs for CATE, CATR, and CATD, and standard errors clustered at the defendant level are used for the 2SLS estimates.

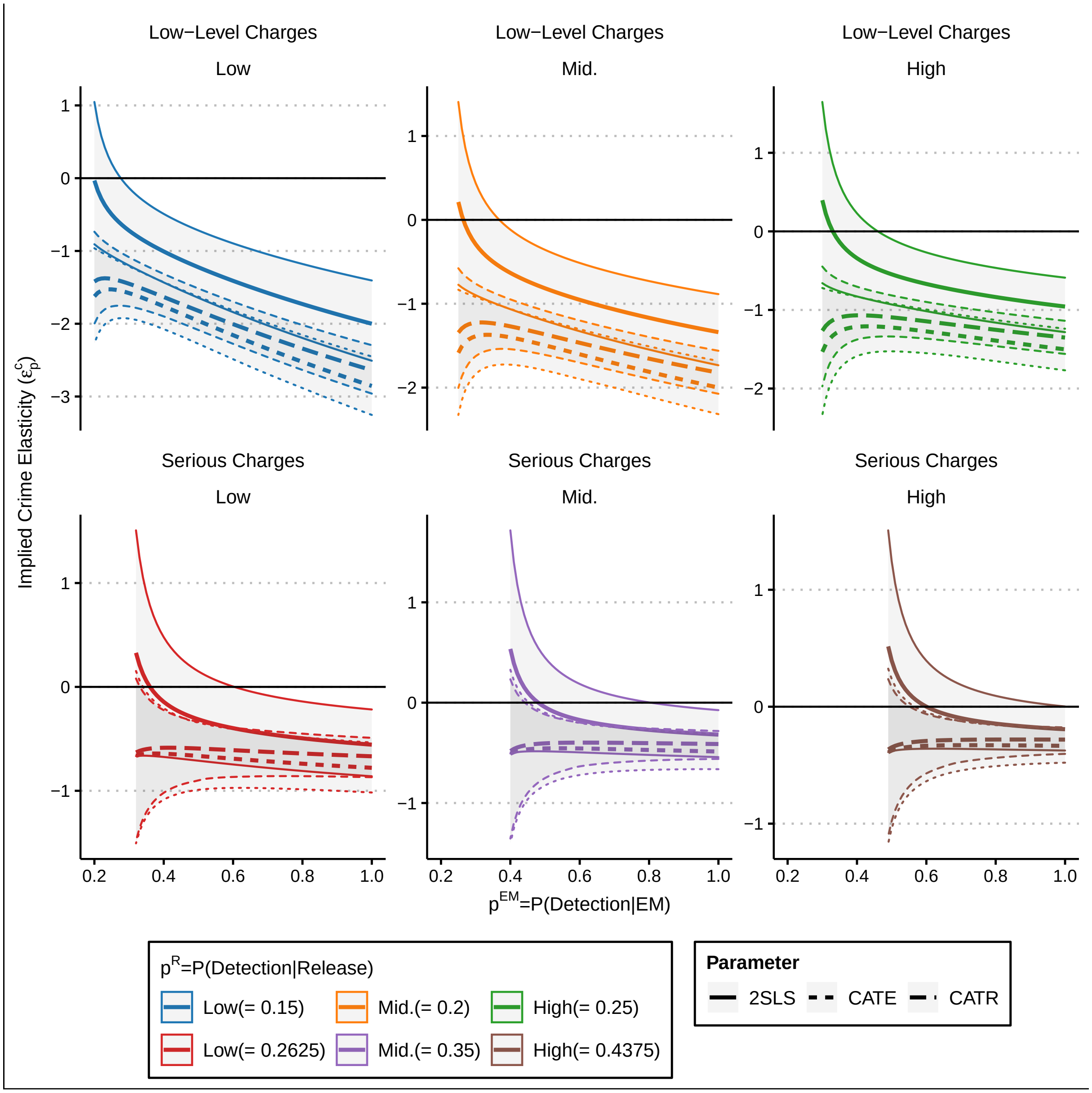


Figure 1.8: Estimated Bounds for Elasticities

Note: Figure displays implied elasticities for different pretrial criminal charge types using a single value for the initial (release) probability of detection across other potential values for the increased probability of detection under EM. Colors correspond to crime type and probabilities of detection under release. Line types correspond to weights being used are for CATE (even weights across defendants and common support) or CATR (higher weights for released defendants across common support). 95% confidence intervals are computed using 200 bootstrap runs or CATE and CATR, and 95% confidence intervals for 2SLS are based on standard errors clustered at the defendant level.

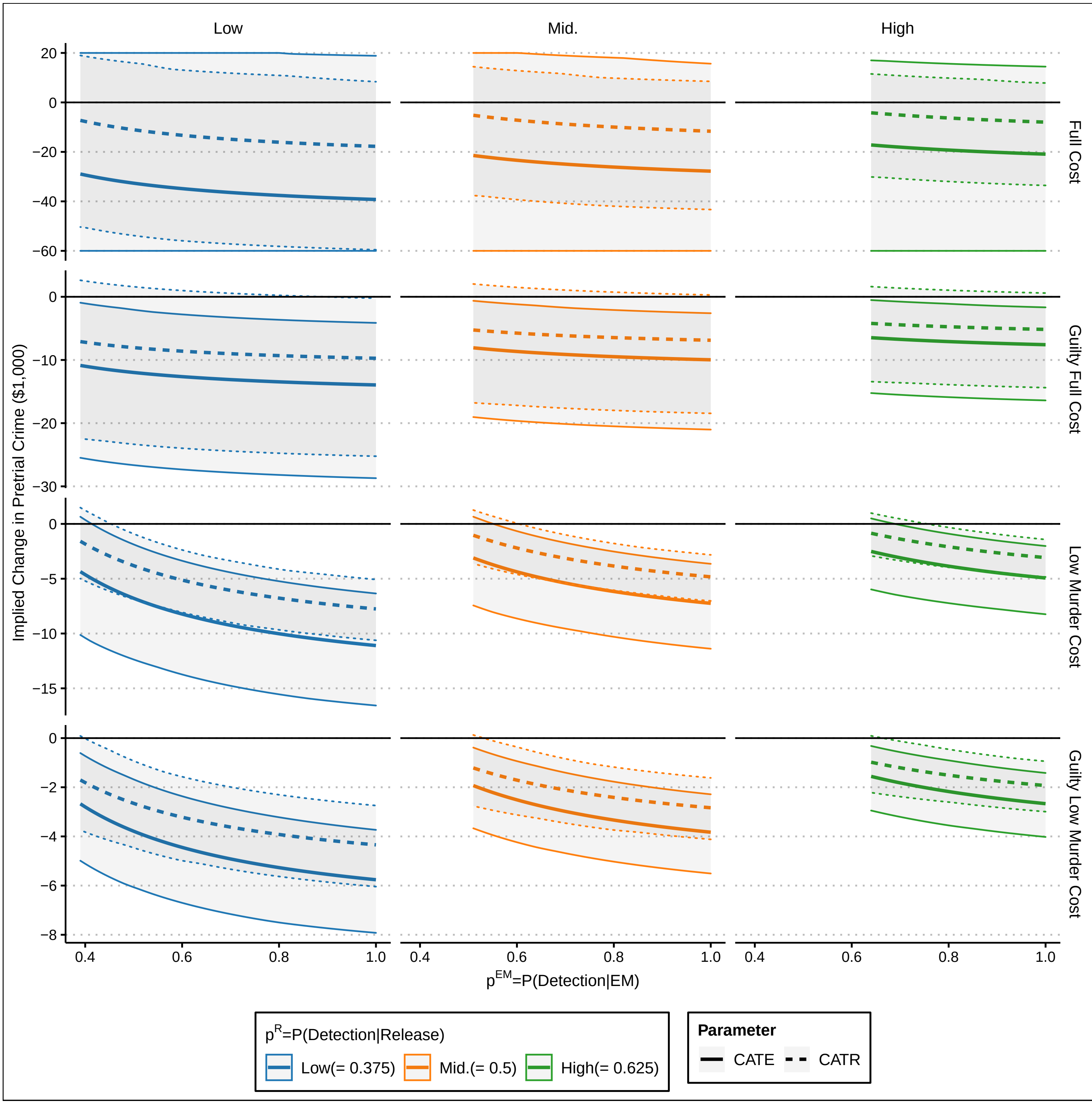


Figure 1.9: Estimated Change in Pretrial Crime Cost: EM vs. Release

Note: Figure displays implied change in pretrial crime costs using a single value for the initial (release) probability of detection across other potential values for the increased probability of detection under EM. Colors correspond to probabilities of detection under release. Guilty charges refers to counting the crime cost only on charges which has a guilty finding. Low murder refers to using a reduced cost of murder in crime cost computations. Line types correspond to weights being used are for CATE (even weights across defendants and common support) or CATR (higher weights for released defendants across common support). 95% confidence intervals are computed using 200 bootstrap runs but are censored in extreme cases for readability.

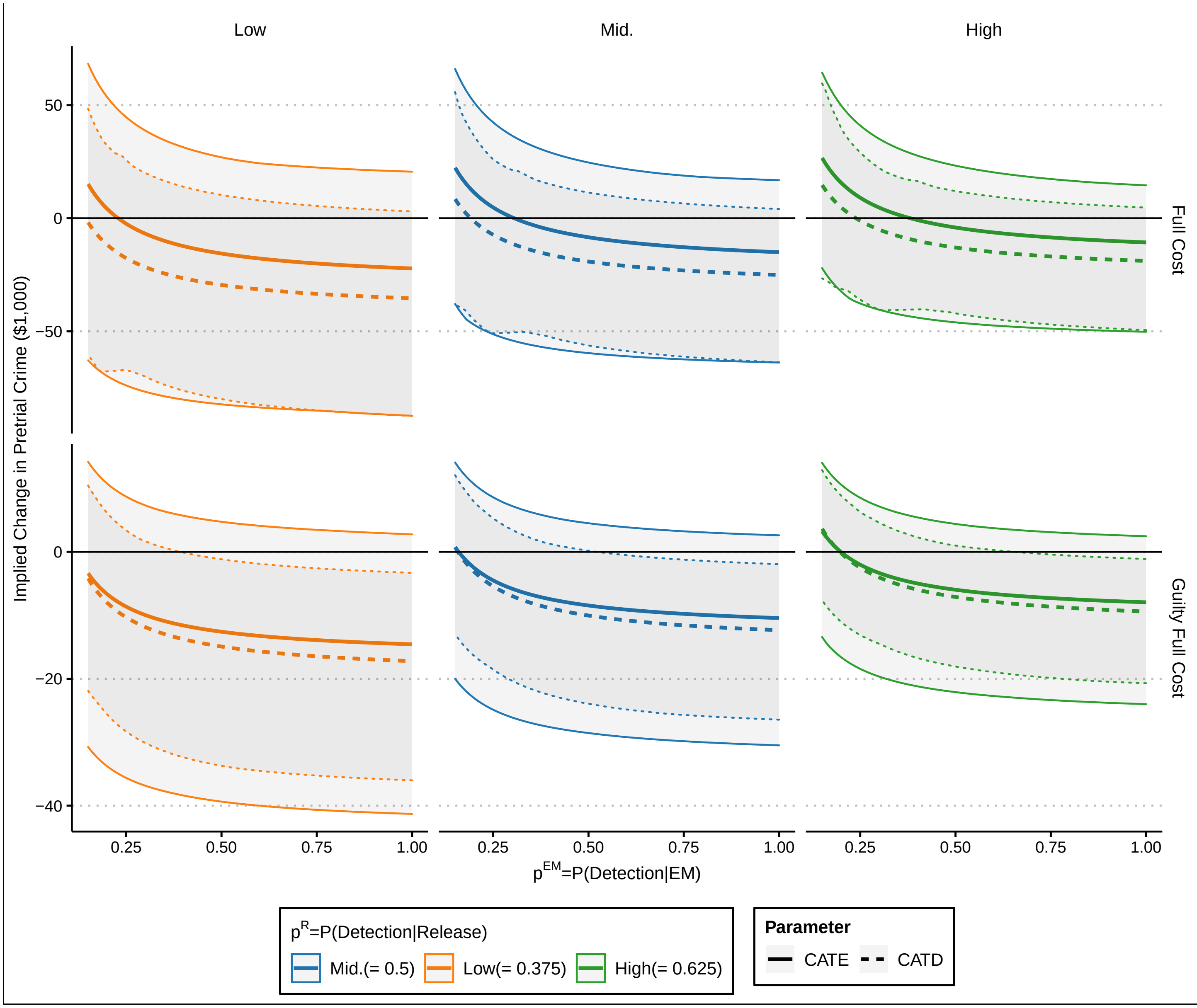


Figure 1.10: Estimated Change in Pretrial Crime Cost: EM vs. Detention

Note: Figure displays implied change in pretrial crime costs using a single value for the initial (release following detention) probability of detection across other potential values for the increased probability of detection under EM. Colors correspond to probabilities of detection under detention. Guilty charges refers to counting the crime cost only on charges which has a guilty finding. Line types correspond to weights being used are for CATE (even weights across defendants and common support) or CATD (higher weights for detained defendants across common support). 95% confidence intervals are computed using 200 bootstrap runs.

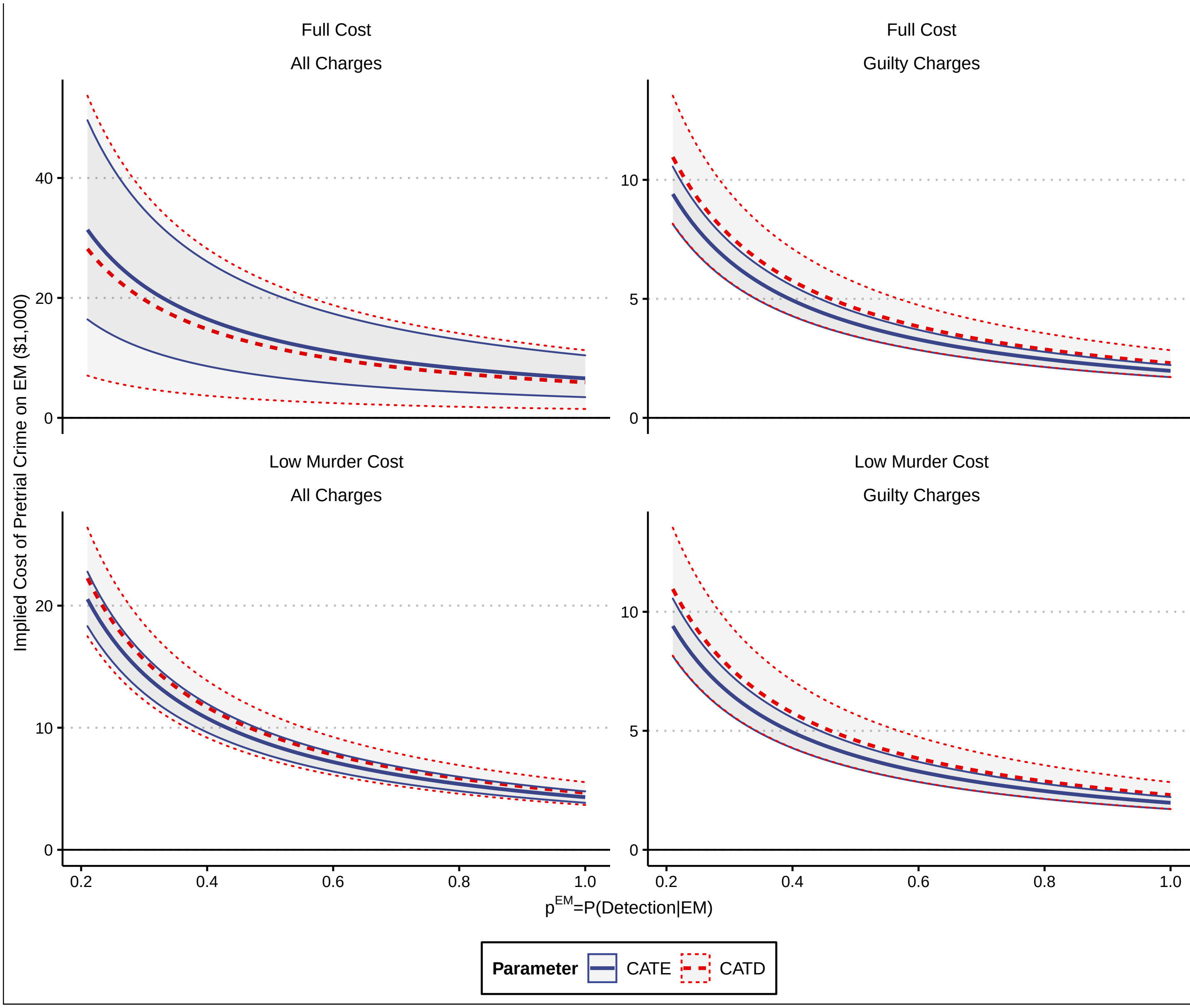


Figure 1.11: Estimated Amount of Pretrial Crime Cost under EM

Note: Figure displays implied amounts of pretrial crime costs for the average defendant and detained defendant across potential values for the probability of detection under EM. Guilty charges refers to counting the crime cost only on charges which has a guilty finding. Low murder refers to using a reduced cost of murder in crime cost computations. Line types correspond to weights being used are for CATE (even weights across defendants and common support) or CATD (higher weights for detained defendants across common support). 95% confidence intervals are computed using 200 bootstrap runs.

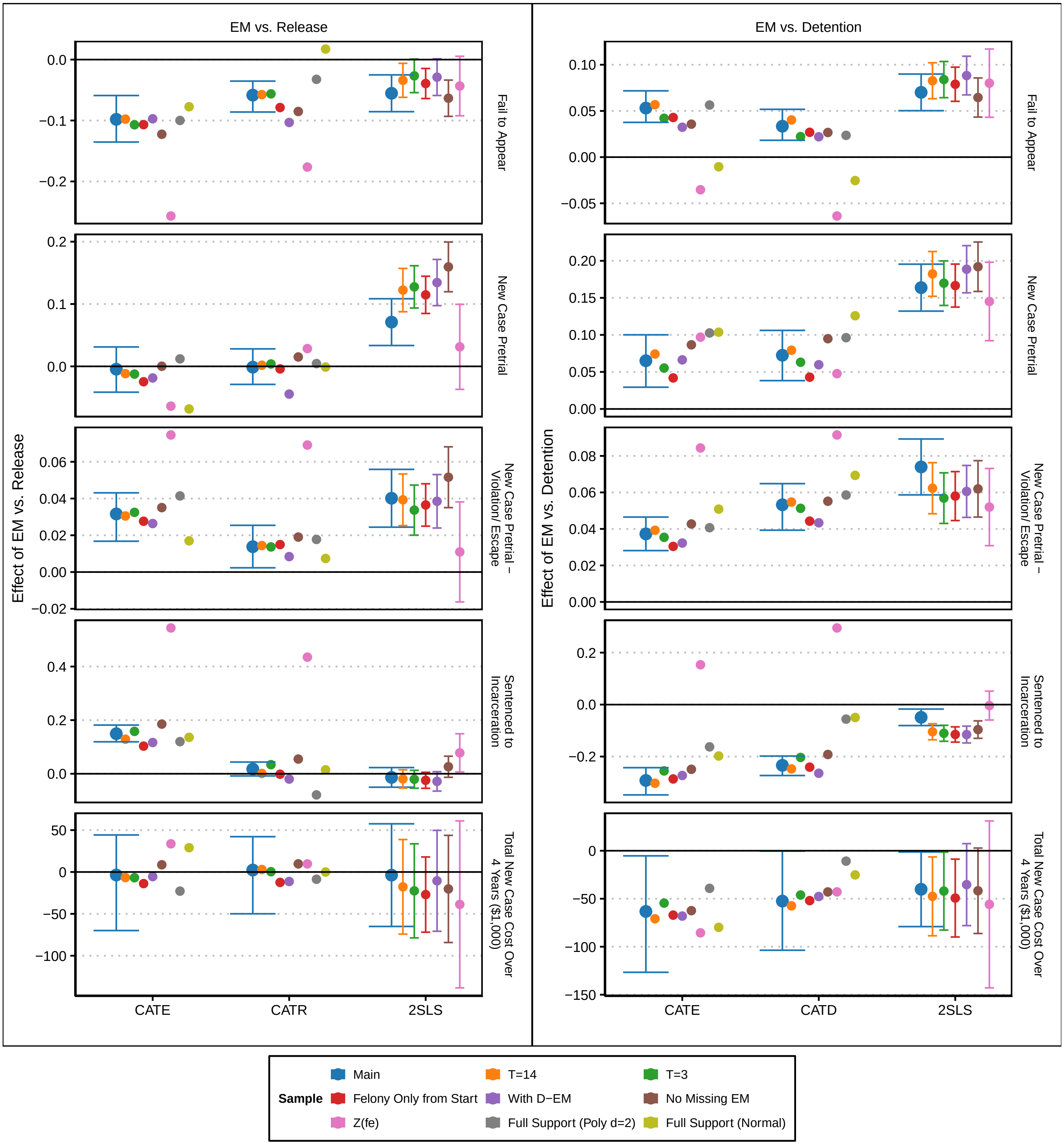


Figure 1.12: Treatment Effects for Main Robustness Tests

Note: Figure displays the CATE, CATR, CATD, and 2SLS estimates for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right). CATE(R/D) are constructed from MTEs that are estimated semiparametrically with equation (6) where Φ_s are 3rd degree polynomials for all samples unless otherwise specified. Poly d=2 (3rd degree polynomial for Φ_s) and Normal (normally distributed errors) are fully parametric and treatment parameters are constructed over full support. 95% confidence intervals of the estimates are computed using 200 bootstrap runs for non-2SLS, while 2SLS confidence intervals are constructed from standard errors clustered at the defendant level.

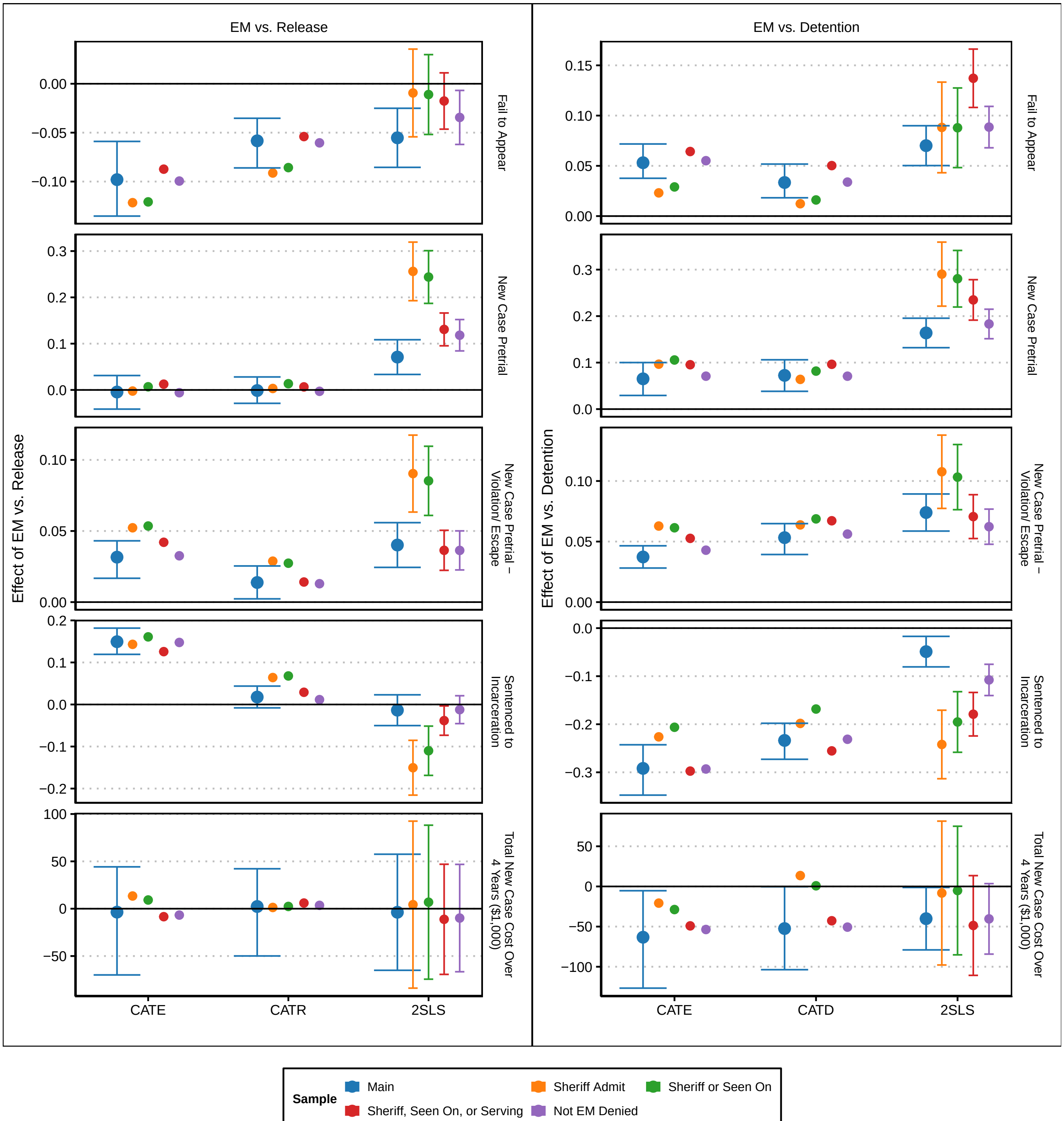


Figure 1.13: Treatment Effects for Recoded Treatments using Disposition Codes

Note: Figure displays the CATE, CATR, CATD, and 2SLS estimates for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right) under different re-codings of pretrial treatments using disposition codes to determine if a defendant was assigned to EM. "Sheriff Admit" means the defendant was explicitly noted to have been admitted into the sheriff's EM program; "Sheriff or Seen On" allows for if the defendant was explicitly noted to be on EM; "Sheriff, Seen On, or Serving" allows for if the defendant was explicitly noted to be serving a monitoring program; and "Not EM Denied" includes "Sheriff, Seen On, or Serving" defendants but excludes any defendant explicitly noted to not be admitted to EM (with bail set to stand). CATE(R/D) are constructed from MTEs recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of the estimates are computed using 200 bootstrap runs for non-2SLS, while 2SLS confidence intervals are constructed from standard errors clustered at the defendant level.

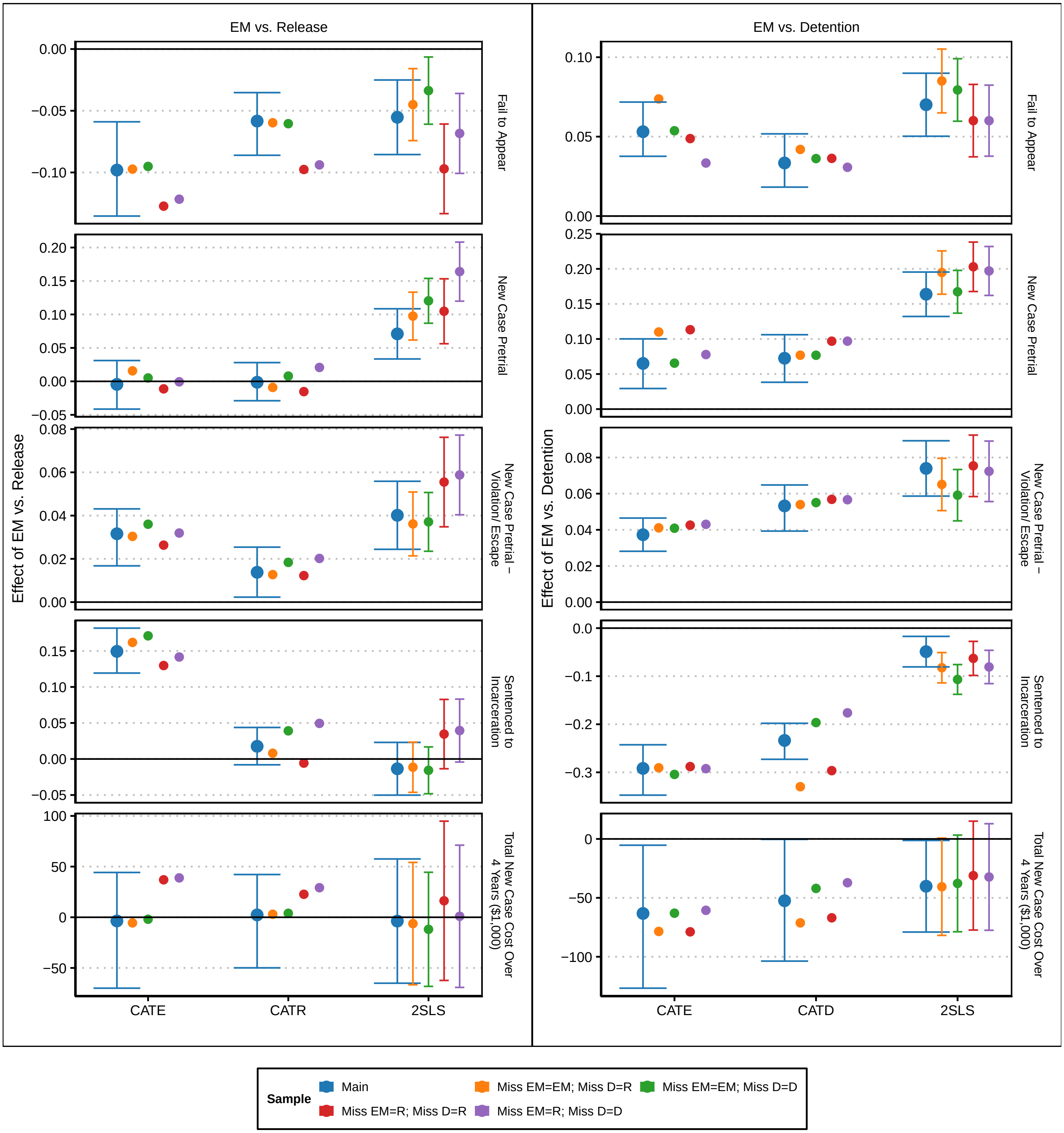


Figure 1.14: Treatment Effects for Recoded Missing Treatments

Note: Figure displays the CATE, CATR, CATD, and 2SLS estimates for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right) under different re-codings of pretrial treatments for cases with missing jail data. CATE(R/D) are constructed from MTEs recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of the estimates are computed using 200 bootstrap runs for non-2SLS, while 2SLS confidence intervals are constructed from standard errors clustered at the defendant level.

	Release (1)	EM (2)	Detain (3)
Median Days in Jail	0	-	58
Median Days on EM	0	41	-
Median Case Duration	33	55	122
Bond Amount	15160.05	26699.44	71615.68
Defendant Demographics			
Def. Black	0.63	0.76	0.77
Def. Hispanic	0.2	0.12	0.12
Def. White	0.17	0.12	0.1
Def. Male	0.85	0.83	0.92
Def Age	33.14	36.75	32.97
Case Characteristics			
Charge - Felony	0.45	0.86	0.77
Charge - Felony Violent	0.01	0.01	0.11
Charge - Felony Drug Poss.	0.28	0.57	0.28
Charge - Felony Drug Deliv.	0.06	0.16	0.14
Charge - Felony Property	0.06	0.13	0.13
Charge - Felony Weapon	0.03	0.01	0.14
Charge - Misdemeannaor	0.58	0.23	0.36
Charge - Misd. Property	0.05	0.02	0.04
Charge - Traffic	0.19	0.16	0.12
Charge - Other	0.06	0.05	0.05
Case History (since 2000)			
Past Cases	5.95	7.86	8.91
Case within Year	0.34	0.41	0.53
Past Failure to Appear	1.3	1.82	2.02
Past Felonies	1.48	2.53	2.69
Past Felonies - Violent	0.1	0.09	0.21
Past Guilty	0.74	1.31	2.07
Past Guilty Felonies	0.26	0.55	0.91
Past Guilty Felonies - Violent	0.02	0.03	0.08
N Obs	34937	22714	26764
Share of Obs	0.41	0.27	0.32

Note: Table displays summary statistics by pretrial treatment for defendants with one observation per defendant and arrest, meaning if there is a felony and a misdemeanor case against a defendant both sets of charges are aggregated to one observation. Variables beginning with 'Charge' are binary variables indicating any charge of a specific type.

Table 1.1: Summary Statistics for Branch 1 Cases by Pretrial Treatment

	Release (1)	EM (2)	Detain (3)
Failure to Appear	0.1420	0.0761	0.0360
Any Guilty Felony Charge	0.221	0.451	0.589
New Case Pretrial	0.145	0.178	0.110
New Case Post-Trial within 4 Years	1.51	1.56	1.67
New Felony Case Post-Trial within 4 Years	0.477	0.616	0.597
New Violent Felony Case Post-Trial within 4 Years	0.0552	0.0427	0.0694

Note: Table displays summary statistics of outcomes by pretrial treatment for defendants with one observation per defendant and arrest, meaning if there is a felony and a misdemeanor case against a defendant both sets of charges are aggregated to one observation.

Table 1.2: Summary Statistics of Outcomes

Est.	(1)	(2)	(3)
LK / LK(1)	1	0.999	0.495
Pseudo-R2	0	0	0.51
Def. Chars		X	
DoW + YM			X
Fes			

Note:
Table displays results from a multinomial logistic regression with the judge assigned to a specific case as the outcome variable to test if defendant observables are predictive of judge assignment. Column (1) is the base model including no regressors, Column (2) only includes case level characteristics, Column (3) contains day of week and year-month fixed effects. LK/LK(1) is the ratio of likelihoods of the model to that of Column (1). Pseudo-R2 is [139]'s measurement of explained variation in the model.

Table 1.3: Testing for Violation of Judge Assignment

	(1)	(2)	(3)
Outcome = Release			
$\hat{Z}^R$ (1 SD)	0.087***	0.087***	0.087***
	(0.00167)	(0.00155)	(0.00152)
Outcome = Detain			
$\hat{Z}^D$ (1 SD)	0.097***	0.097***	0.097***
	(0.00152)	(0.00139)	(0.00136)
Controls		X	X
Month + DoW FEs			X

Note: Table displays the relevance of the standardized instruments in predicting whether or not the defendant is assigned to Release or Detention. Column (1) contains no controls, Column (2) adds case and defendant controls, and Column (3) adds time fixed effects. Standard errors are clustered at the defendant level.

Table 1.4: Relevance Tests

	EM vs. Release	EM vs. Detain
	(1)	(2)
Pretrial Misconduct		
Fail to Appear	-0.055***	0.07***
	(0.015)	(0.01)
New Case Pretrial	0.071***	0.16***
	(0.019)	(0.016)
New Case Pretrial - Serious	0.016	0.056***
	(0.012)	(0.011)
New Case Pretrial - Low-Level	0.035**	0.086***
	(0.017)	(0.014)
New Case Pretrial - Violation/ Escape	0.04***	0.074***
	(0.008)	(0.0078)
Case Outcomes		
Any Guilty Felony Charge	0.027	-0.016
	(0.021)	(0.014)
Sentenced to Incarceration	-0.014	-0.049***
	(0.019)	(0.016)
Post-Trial New Cases		
Total New Cases Post Trial within 4 Years	0.16	0.098
	(0.11)	(0.1)
Total New Cases Post Trial within 4 Years - Felony	0.017	-0.016
	(0.053)	(0.044)
Total Case Cost (Pre and Post) ($1,000)		
Total New Case Cost Over 4 Years	-3.7	-40**
	(31)	(20)
Total New Case Cost Over 4 Years (Low Murder)	0.12	-0.071
	(2.6)	(1.9)
Total New Case Pretrial Cost	-18*	-8.3
	(9.5)	(6.6)
Total New Case Pretrial Cost (Low Murder)	0.15	2.6***
	(0.93)	(0.72)
Total New Case Post-Trial Over 3 Years Cost	20	-17
	(26)	(15)
Total New Case Post-Trial Over 3 Years Cost (Low Murder)	-0.03	-1.8
	(2.1)	(1.5)
Min. N	51,327	51,327

Note: Table displays the results of 2SLS regressions for the main felony sample for the effect of EM vs. Release and EM vs. Detention. Includes case level controls and quarter and weekday fixed effects. Standard errors clustered at the defendant level are in parantheses. ***p < 0.01; **p < 0.05; *p < 0.1

Table 1.5: 2SLS Results for Main Sample

	EM vs. Release		EM vs. Detain	
	CATE	CATR	CATE	CATD
	(1)	(2)	(3)	(4)
Pretrial Misconduct				
Fail to Appear	-0.0981***	-0.0583***	0.0531***	0.0334***
	[-0.135 , -0.059]	[-0.0861 , -0.0353]	[0.0376 , 0.0718]	[0.0182 , 0.0518]
New Case Pretrial	-0.0045	-0.0013	0.0651***	0.0725***
	[-0.0414 , 0.0311]	[-0.0289 , 0.028]	[0.0294 , 0.1]	[0.0383 , 0.106]
Serious Charge	-0.003	-0.0035	0.0002	0.0035
	[-0.0263 , 0.0166]	[-0.0244 , 0.0126]	[-0.0228 , 0.0229]	[-0.0187 , 0.0243]
Low-Level Charge	-0.0298*	-0.0185	0.0382**	0.0307*
	[-0.066 , 0.0019]	[-0.0443 , 0.0105]	[0.0071 , 0.0661]	[-0.0009 , 0.0576]
Violation/ Escape Charge	0.0316***	0.0138**	0.0373***	0.0532***
	[0.0168 , 0.0431]	[0.0023 , 0.0254]	[0.0281 , 0.0465]	[0.0393 , 0.0648]
Total Cost of Pretrial New Cases ($1,000)				
Total Pre (Full Murder Cost)	-10.6	-2.47	-4.2	-9.57
	[-40.9 , 10.7]	[-18.8 , 7.34]	[-29.8 , 12.3]	[-27.8 , 5.7]
Total Pre (Low Murder Cost)	-1.47	-0.442	-0.465	-0.851
	[-3.63 , 0.416]	[-1.74 , 0.712]	[-2.6 , 1.63]	[-2.7 , 0.661]
Case Outcomes				
Any Guilty Felony Charge	0.0607**	-0.0278	-0.305***	-0.222***
	[0.0145 , 0.112]	[-0.0688 , 0.0098]	[-0.359 , -0.25]	[-0.266 , -0.186]
Sentenced to Incarceration	0.149***	0.0176	-0.292***	-0.234***
	[0.119 , 0.182]	[-0.008 , 0.0438]	[-0.347 , -0.243]	[-0.273 , -0.198]
Total Post-Trial New Cases within 4 Years				
Total Cases Post within 4 Years	0.389***	0.351***	-0.0166	0.281***
	[0.207 , 0.593]	[0.202 , 0.501]	[-0.239 , 0.245]	[0.111 , 0.468]
Total Felony Cases Post within 4 Years	0.134***	0.106***	0.01	0.143***
	[0.0364 , 0.216]	[0.033 , 0.184]	[-0.084 , 0.118]	[0.067 , 0.234]
Total Cost of Post-Trial New Cases within 3 Years ($1,000)				
Total Post (Full Murder Cost)	6.04	7.48	-44.4*	-31.2
	[-60.8 , 41.8]	[-36.6 , 45.5]	[-84.2 , 0.414]	[-72.3 , 10.7]
Total Post (Low Murder Cost)	2.86	3.02**	-2.28	0.704
	[-1.34 , 6.42]	[0.0141 , 5.94]	[-5.39 , 1.73]	[-2.64 , 4.22]
Total Cost of New Cases within 4 Years of Bond Court ($1,000)				
Total Pre and Post (Full Murder Cost)	-3.62	2.32	-63.2***	-52.4**
	[-69.9 , 44.2]	[-49.9 , 42.2]	[-127 , -5.32]	[-104 , -0.337]
Total Pre and Post (Low Murder Cost)	2.31	2.32	-5.73**	-1.41
	[-2.45 , 7.08]	[-1.05 , 5.77]	[-11.1 , -0.0579]	[-5.55 , 2.89]

Note: Table displays the common support average treatment effect (CATE) and average treatment effect on the released and detained (CATR, CATD) estimates averaging over the MTEs. 95% confidence intervals are displayed below the estimate. 95% CIs and p-values are computed through 200 bootstrap runs, with each run calculating the effect, then the lower bound and upper bound are the bootstrap treatment effects at the 1st, 2.5th, 5th, 95th, 97.5th, and 99th percentile. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. ***p < 0.01; **p < 0.05; *p < 0.1

Table 1.6: ATE, ATR, and ATD Results for Common Support

	EM vs. Release		EM vs. Detention	
	ATE	ATR	ATE	ATD
(Direct) Jail/EM	0.44	0.23	-7.19	-8.96
	[0.4 , 0.47]	[0.2 , 0.26]	[-8.17 , -6.46]	[-9.46 , -8.54]
Failure to Appear	-0.1	-0.06	0.05	0.03
	[-0.14 , -0.06]	[-0.09 , -0.04]	[0.04 , 0.07]	[0.02 , 0.05]
Pre-Trial Crime	-46.31	-19.41	-24.99	-41.76
	[-147.06 , 26.1]	[-72.36 , 14.17]	[-106.29 , 28.05]	[-105.04 , 6.83]
Pre-Trial Crime (Low Murder)	-12.08	-8.05	-8.74	-10.62
	[-18.98 , -6.07]	[-11.67 , -4.67]	[-15.38 , -1.99]	[-16.12 , -6.14]
Sentencing	2.99	0.35	-5.84	-4.67
	[2.39 , 3.63]	[-0.16 , 0.88]	[-6.95 , -4.85]	[-5.46 , -3.96]
Post-Trial Crime (3 Years)	20.12	24.94	-147.9	-103.93
	[-202.7 , 139.37]	[-121.93 , 151.62]	[-280.61 , 1.38]	[-240.83 , 35.6]
Post-Trial Crime (3 Years) (Low Murder)	9.52	10.05	-7.6	2.35
	[-4.47 , 21.39]	[0.05 , 19.79]	[-17.97 , 5.78]	[-8.79 , 14.07]
Total	-22.86	6.06	-185.87	-159.3
	[-347.11 , 169.51]	[-194.34 , 166.9]	[-401.99 , 18.18]	[-360.77 , 29.99]
Total (Low Murder)	0.77	2.53	-29.31	-21.88
	[-20.79 , 19.36]	[-11.67 , 16.22]	[-48.44 , -7.46]	[-39.81 , -4.51]

Note: Table displays the relative costs (positive) and savings (negative) per defendant under counterfactual policies, where the average defendant is moved from release to EM or detention to EM (ATE), where the average released or detained defendant is moved to EM (ATR or ATD). Estimates are constructed using common support analogs of MTRs with costs or scaling applied and weighted by the relevant treatment parameter weights. I use average costs of $15 per defendant-day for EM and $150 defendant-day for jail and apply a 40% marginal cost. FTAs are valued at $1,000. Crime costs are based on aggregated arrest costs assuming a 30% cost-weighted detection rate under release, post-detention release, or post-trial, and a 60% cost-weighted detection rate under EM. Pretrial crime costs include punishment costs. Sentencing costs are assumed to be $20,000 per defendant sentenced to incarceration, based on [137] estimates of average prison cost per defendant in Illinois in 2015 of $33,507 and applying a 60% marginal cost discount. This also assumes defendants sentenced to incarceration are serving one year on average, which may be an underestimate (over 75% of individuals sentenced to prison or jail are sentenced to prison). 95% confidence intervals are displayed below the estimate. 95% CIs are computed through 200 bootstrap runs.

Table 1.7: Breakdown of Relative Counterfactual Costs ($1,000)

Chapter 2: The Effect of Minority Peers on Future Arrest Quantity and Quality

Roman Gabriel Rivera

2.1 Introduction

Aggressive policing, such as the excessive use of force and over-policing of low-level crimes, has numerous social and economic costs and disproportionately affects minority communities. Increasing minority representation in policing is one of the most common policy proposals to address aggressive policing, as well as to build community trust and to improve police legitimacy ([140]) Such policies address the disproportionately white and male composition of police departments relative to the communities they police, as highlighted in [141] who finds a 24 percentage point gap in minority representation. Existing research finds that minority officers police minority civilians less aggressively ([142], [143], [144]). However, the effectiveness of these policies depends on both the direct effect of these officers as well as the spillover effects of increased diversity. Minority officers may influence their peers' policing in myriad ways: interracial friendships may reduce racial bias; negative interactions can lead to animus and prejudice; or officer's may adopt their peers' preferences for policing behavior, such as aggressive policing. Thus, while the full effect of increasing diversity in policing hinges on which of these mechanisms is at play, we still know very little about how peer diversity affects officer behavior.

In this paper, I provide novel evidence of increased shares of minority officers reducing their peers' aggressive policing and improving their peers' arrest quality. This indicates that increased diversity has positive spillovers in policing. Such evidence has previously proved elusive for two main reasons. First, it requires peer groups to be quasi-randomly assigned to avoid self-selection.

Second, it requires data on individual police officer demographics, arrests and arrest quality, and peer groups, that are difficult to obtain. I overcome both of these obstacles using detailed data on police officers in the Chicago Police Department (CPD) who were randomly assigned to peer groups in the form of police academy cohorts based on lottery numbers. The police academy is a highly relevant environment: increased departmental diversity requires training increasingly diverse cohorts; the academy forms recruits' first major experiences and relationships in policing; and the CPD academy involves training with one's cohort for 6 months (900 hours), making the peer diversity treatment significantly more intense than most diversity focused interventions.

I first document that officers assigned to police academy cohorts with higher shares of minorities (Black, Hispanic, Asian, and Native American) make fewer arrests of minorities (which represent > 90% of arrests in Chicago) once they become full officers. However, I find that cohort composition also has a minor influence on where officers work post-academy, which influences arrest opportunities. To correct for this, I control extensively for officers' working conditions and recover individual officer propensities to make arrests of various types using a novel data set on millions of daily shifts and assignments. Consistent with the previous result, higher shares of minorities in academy cohorts reduce officers' future propensities to arrest minorities, especially Black civilians, even after taking working environment into account.[1] Notably, I do not only study the effect of minority peers on whites, as is common in the peer diversity literature. Rather, I study the effect of minority peers on all officers to understand the full extent of spillovers, as minorities make up sizable shares of police departments.

A decline in arrests may be due to either decreased public safety or a reduction in potentially harmful and overly aggressive policing. To distinguish between these two possibilities, I dissaggregate arrests into arrests for serious (e.g., violent, property) and low-level (e.g., drug, traffic) crimes. As serious arrests are crucial for maintaining public safety, a reduction is generally unde-

[1]It is unsurprising that arrests of Black civilians are driving the results, as they make up the vast majority of new officer arrests (> 80%) and differences in enforcement activity between officers tend to be most salient in enforcement against Blacks even among non-Black officers (e.g., male and female, white and Hispanic) ([144]). However, minority peers have a negative effect on low-level arrests of all groups (e.g., white civilians), though not all point estimates are precisely estimated.

sirable. However, there is mounting evidence that arrests and prosecution of low-level crimes can significantly harm individuals and actually lead to future criminality ([33]).[2] I find that minority peers have a large negative effect on an officer's propensity to make low-level arrests while having a small positive or null effect on serious arrests. For example, a 5pp increase in Black peers ($\approx$ 1 SD) decreases an officer's propensity to arrest Blacks for low-level crimes by 0.16 standard deviations. Furthermore, this effect is enduring as it is present even four years after the academy has ended.

Beyond the type of an arrest, measuring the 'quality' of an arrest is important as well. If minority peers cause large reductions in productive arrests, then we can interpret the effect as a reduction in the quality or effort of officers. To measure arrest quality, I link arrest records to court outcomes and recover officers' propensities for high (guilty finding in court) and low (no guilty finding) quality arrests for serious and low-level crimes. Prior work measuring arrest quality is limited due to the difficulty in obtaining and linking detailed arrest and court records.[3] I find that the decline in propensities to arrest Blacks for low-level crimes is driven almost entirely by a decline in low quality (not found guilty) arrests with little to no effect on high quality (found guilty) arrests. Combined with the small and positive effects on serious arrests, this indicates that minority peers reduce aggressive policing (low-quality low-level arrests) and increase average arrest quality, while not negatively affecting officer effort towards serious crime.

Understanding the mechanisms driving these effects is crucial to develop appropriate policy. As mentioned above, there are multiple potential effects, and I find evidence consistent with two. First, I find that minority peers cause the largest decreases in aggressive policing in white officers. This is consistent with positive interracial contact between whites and minorities reducing bias against Blacks, leading to less aggressive policing. As expected, Black peers have the largest effect on whites with respect to low-level arrests of Black civilians, but other minority peers (Hispanics and other non-whites) also reduce whites' aggressive policing of Blacks. This latter effect may be

[2] See also [37], [28], [31], and [30].

[3] [145] also uses court outcomes as a metric for arrest quality. However, my data allow me to estimate an individual officer's propensity to make high and low quality arrests within race and crime-type. [146] also discusses arrest quality, but only use whether the arrest led to a charge as a measure of quality.

a result of non-Black minorities facilitating Black-white socialization or positive contact between whites and other minorities indirectly reducing anti-Black bias.[4]

Second, while peer diversity studies often conceptualize peer effects as operating in one direction by examining the effect of minorities (the treatment) on whites (the unit of study), there is evidence that all peers influence each other's actions through forming group cultures, influencing social identity, and altering each others' preferences and personalities ([150], [151], [152], [153]). I find evidence consistent with this. All officers, white and minority, are influenced more strongly by peers whose observable characteristics are associated with less aggressive policing. This is consistent with officers assigned to cohorts with peers who have lower propensities for aggressive policing also policing less aggressively in the future. Based on the magnitudes, the results suggest that the effect on all officers drives the results and is larger than the whites-only effect.

This paper builds on the literature on diversity and policing. Prior research finds that minority and female representation influences city-level policing outcomes ([154], [155]), and individual officer race and gender are associated with differential policing behavior ([143], [144]).[5] This paper extends this research by identifying peer diversity as a causal determinant of officer's future policing behavior. One implication is that the changes in department-level outcomes resulting from diversity initiatives were the result of both the direct effect of more diverse officers and the indirect effect on their peers. I also build on this literature by studying how peer diversity affects officers' arrest quality, a new outcome in this literature, in addition to their arrest quantity.

This paper also contributes to the growing literature on officer-level interventions. For example, [161] find that officers assigned to meetings with supervisors were 12% less likely to make an arrest. By comparison, I find a similar reduction can be achieved for low-level Black arrests through a 6.2 percentage point increase in older minority peers in the police academy.[6]

---

[4]Positive contacts with one minority group can cause spillovers of improved sentiment toward other minority groups through the secondary transfer effect ([147], [148], [149]).

[5]Notably, [154] no effect of affirmative action litigation on crime, but does find a decrease in Black arrest share among serious arrests, where as this paper documents larger peer effects on low-level, not serious, Black arrests. For the effect of representation on more macro-level outcomes see [156], [157], and [158], and [159]. For the more on the relationship between officer race and policing outcomes see [160] and [142].

[6]Older being defined as > 32 years, and relative to the share of young (< 27 year) white recruits.

By providing evidence for two main mechanisms, I also contribute to the literatures on both interracial peer effects and the effects of social identity on behavior.[7] First, a substantial literature in psychology and economics studies how minorities, often Blacks, influence the perceptions, biases, and beliefs of whites. They generally find that increased interracial socialization with minorities improves whites' perceptions of minorities ([173]), increases openness to future contact with minorities ([174]), reduces anti-minority decisions ([175]), and reduces implicit bias and decreases participation in racist politics across generations ([176]).[8] In line with this, I find that minority peers cause larger decreases in aggressive policing of Black civilians among white officers relative to minority officers.

Second, peer diversity studies generally focus solely on the effect of minorities on whites. I expand on this by studying the peer effect of diversity on minorities as well. A smaller literature studies how peers influence outcomes through shifts in preferences and social identity.[9] For example, [152] find that jurors' political alignment influences trial outcomes through changing peer opinions. I provide multiple findings consistent with the effect of peer preferences on officer behavior, most centrally that minority peers influence both white and minority officers.

This paper proceeds as follows. In Section 2.2, I describe the background and data for this paper. In Section 2.3, I discuss the empirical strategy, and Section 2.4 presents the results. Section 2.5 contains robustness checks, and Section 2.6 explores mechanisms and alternative explanations. Section 2.7 concludes.

[7] I employ a common identification strategy (random assignment of students to classrooms) in the educational peer effects literature ([162]) but in a new setting. See [163] and [164]. [165] discusses various studies in the educational peer effects literature. [166] studies police academy cohorts as well, but their identification hinges on a difference-in-differences design, similar to [167] which studies pilots. [168], [169], [170], [171], and [172] provide evidence for peer composition in educational environments influencing future outcomes.

[8] See also [177], [178], [179], [147], [180], [181], [182], and [183].

[9] See [150], [184], [185], [186], [153], and [187].

2.2 Background and Data

2.2.1 Chicago Police Department and Recruitment

Application to CPD and the Academy

Comprised of over 10,000 officers, the Chicago Police Department (CPD) is the second largest police force in the US. It polices the nation's third largest city, which is racially diverse and economically segregated. To recruit new officers, the CPD issues a call for officers, and applicants take a written exam, which they must pass in order to enter the academy. As a CPD Frequently Asked Question (FAQ) form, [188], explains:

> All applicants who pass the exam are placed on an eligibility list based on a randomly assigned lottery number. You will be referred to the Chicago Police Department in lottery order as vacancies become available.

After an applicant's number is called, if they pass required physical and psychological tests, they are permitted to start at the police academy (see Appendix B.1.1 for more discussion). Academy start dates, known as "appointed dates", correspond to officers beginning their time at the police academy. In Appendix B.1.2, I provide empirical support for the random assignment of officers to cohorts. I define a cohort as the group of officers with the same appointed date—in the main sample, cohorts are separated by about 1 month. During the academy, officers must complete 900 hours (about 6 months) of training in multiple areas, such as "firearms, control tactics, physical training, [and] classroom training" ([189]).[10]

After the academy, the recruits in a cohort enter an on-the-job-training period for one year as "probationary police officers" during which they are split up, work in multiple areas of the city, and are evaluated under the supervision of Field Training Officers. After recruits meet the various requirements, complete their time as a probationary officers, and become "field qualified" ([190]),

[10]The 900 hours of training encompasses and surpasses the training required to pass the Illinois State Peace Officer's Certification Exam. In larger cohorts, officers are further subdivided into "homerooms" which take most of their trainings together– while data on homerooms could not be obtained, I use detailed training data to approximate these groups.

they exit their probationary period and become a full (sworn) Chicago police officer. New sworn officers are then assigned to more permanent units.

Sample and Requirements

This paper focuses on the cohorts with start dates between 2009 and 2016, as they overlap best with the data. These cohorts can be divided into three periods based on the year during which officers took the entrance exam, i.e. the level at which they were randomly assigned lottery numbers. (See Figure B.2.1 for exam information.) In 2006, three exams were given in rapid succession, each attended by a relatively small number of applicants (all between 800 and 1,500 passing applicants)– these tests will be collectively referred to as Exam 2006. The next test was issued in 2010 (Exam 2010), with almost 8,000 passing applicants. In 2013, the final test in our sample was issued (Exam 2013) with over 12,000 passing applicants and an important policy change: the minimum age of entry was reduced from 23 to 21 ([191])– the maximum age is 40 years old for all cohorts in the sample. While the CPD did not provide information on which officers belong to which test, in Appendix B.1.1 I show that the Exam 2010 cohorts can be identified with near certainty.

The CPD is massively over-subscribed: fewer than 3,000 applicants were called into the academy between 2009 and 2016, while over 20,000 applicants passed the exams. This is because CPD jobs are highly desirable by individuals from across the country, and applicants to law enforcement are highly passionate about joining a police force. In this sample, the CPD began to call individuals into the academy years after their respective exam was taken: the last batch of 2006 test-takers were likely called between March 2009 and October 2011; 2010 test-takers were first called into the academy between April 2012 and May 2014; and 2013 test takes were likely called between August 2014 and December 2016.[11]

[11] Based on internet discussions, passing applicants with a high lottery number (far into the queue) are advised that "They will probably test again before they call you" and that "It may take another year or so but you will keep moving up the list... Just keep training and stay strong and good luck!!!" ([192]).

Units and Daily Assignments

Transfers between assignments and the filling of vacancies are determined by a seniority-based bidding process and are only available to non-probationary sworn officers, meaning new officers have little to no choice in where and when they work ([193]). New officers are generally assigned to the patrol units which correspond to geographical districts in Chicago.[12] These units occupy most CPD officers and relate to what is commonly considered police work. There are many other units for specialized work that contain far fewer and more experienced officers, such as training units, detective units, etc., which are not studied in this paper.

Within units, officers also bid for shifts/watches (generally, 12 am - 8 am, 8 am - 4 pm, 4pm - 12 am), their 'day off group', and furlough days at the end of the preceding year– this is also seniority based. On any specific day, whether or not an officer is assigned to work depends on their rotating schedule, which is generally 4 days on and 2 days off during the period of study, and is predetermined by their day off group based on the CPD operations calendar (see Figure B.2.2 and [144] for more details). This means that the exact days an officer works are predetermined, not up to the officer's discretion on that day, and rotate over the days of the week.[13] Furthermore, it means that the exact composition of the officers working in a unit and watch on a specific day will be different the following week as officers of different day of groups and furlough schedules will be working together– officers do not frequently work in different shifts throughout a year.

2.2.2 Data

The data for this study come from the Chicago Police Department, Chicago's Department of Human Resources, and the Circuit Court of Cook County (in which Chicago is located). By combining data sets on CPD officers obtained over five years, I construct a detailed panel data set of officer assignments, arrests, and arrest outcomes in court between 2010 and 2018. This contains

[12]There were 25 districts / patrol units before 2012. During 2012, three of these units/districts were collapsed into other districts, reducing the total number to 22.

[13]For example, one week an officer works Tuesday through Friday, and the next week they work Monday through Thursday

officers' demographic information (race, gender, age), start dates, when officers exited the training unit (after the academy and probationary period), and other administrative information. Daily assignment and attendance data includes daily records on officer assignments and time on duty for the geographic units. Additional data sets contain information on trainings, officer education, military status, and language ability. Collectively, these data permit highly granular analysis of an officer's working environment and peer groups. I restrict my analysis to observations of police officers (the lowest and most common rank, e.g., not detectives, sergeants, etc.) working on shift (watch) numbers 1-4, and assigned to regular assignments (e.g., not administrative, lockup, desk duty, etc.).

In order to recover individual officer's arrest quantity and quality metrics, I use arrest data and court data. The arrest data contain all arrests of adults by Chicago police officers including arrest date and time, crime description, primary arresting officer(s), and arrestee race.[14] By linking the arrest data to court records, I construct a metric for arrest quality by determining if the arrest was associated with any guilty finding, which indicates high quality, or no guilty, which indicates low quality. Guilty findings include plea deals, which account for over 90% of convictions. Combined, these data allow me to construct a measure for individual officer's arrest quantity and quality after extensively controlling for their working environments. For more discussion of the data, see Appendix B.3.

Sample Selection

A total of 2,795 officers joined the CPD between March 2009 and December 2016. As defined above, an academy cohort is all the recruits who started at the CPD academy on the same date, resulting in 69 cohorts during this period. I focus my primary analyses on the Exam 2010 cohorts (the "Main Sample"), starting between July 2012 and May 2014, for multiple reasons. First, I can observe main sample officer assignments and arrests from their probationary periods onward, and during their time at the academy (July of 2012 to mid-2015), there were no changes in departmental

[14] Almost all arrests have at most two primary officers listed. Data on juvenile arrests is much more limited as it is protected from FOIA; for example, central booking number, race, gender, and crime type are redacted.

leadership or major political or policing scandals in Chicago. Relative to the other exams, the assignment patterns of Exam 2010 cohorts are most consistent the random assignment assumption (see Appendix B.1.2), and main sample cohorts originated from the same entrance exam issued in December of 2010 (see Appendix B.1.1), which is not as certain for the 2006 test or 2013 test cohorts. Furthermore, there are twice as many officers in Exam 2010 relative to Exam 2006, and Exam 2010 officers can be observed twice as long into their careers relative to Exam 2013 officers. Lastly, I am able to include data on field training officers for Exam 2010, but not the other cohorts. The downsides to using only the main sample are that there will be fewer cohorts and that cohorts will be the only level of variation. I refer to Exams 2006, 2010, and 2013 collectively as the "Full Sample".

All officers in the full sample were subject to a series of filters.[15] Notably, I drop recruits in cohorts who were not matched in the assignment data, recruits with invalid durations in the academy or probationary period, and recruits not matched in the salary and unit assignment data. I also drop a few recruits that had fewer than 15 observations in the assignment panel. Attrition from the initial cohort to the final sample can occur for multiple reasons. If attrition is related to cohort composition, it may contaminate the results. But, as I show in Appendix B.1.3, cohort diversity has no significant impact on attrition in the main or full samples. After filters, the main sample of cohorts contain 940 new officers in 21 cohorts with 531,597 total officer-shift observations over 61 months. The full sample, likewise, contains 2,336 officers in 43 cohorts with 1,081,543 total officer-shift observations over 100 months.

[15]Before calculating initial cohort composition, I excluded very small cohorts that started during the sample period but had cohorts with fewer than 7 recruits which removed a total of 37– the majority were in single officer cohorts. These small 'cohorts' are likely errors as the next smallest cohort size is 25. I also dropped 1 recruit who reported starting too young, and another 9 officers who likely had erroneous start dates were also removed.

2.2.3 Summary Statistics

Cohort Composition

Table 2.1 displays the demographic composition of each exam period (Exam 2010, 2006, 2013) in Columns (1)-(6) with even columns containing pooled means and odd columns containing means over cohort compositions before attrition. Column (7) contains the pooled demographics of all officers in the panel data for reference. By comparing Columns (1) and (2), it is apparent that the main sample of recruits is very similar to that of their average cohort, which is expected due to the random assignment of recruits to cohorts, and that attrition after entering the academy did not significantly alter the demographic composition of the pool of officers. This similarity is also apparent by comparing the pooled and cohort means for the other exam periods. Overall, the main sample of recruits is 80.74% male, start at 30.1 years old, and 48.72% minority– which is comprised of mostly Hispanics (31.17%) and Blacks (13.19%). The average main sample cohort contains 54 recruits.

The comparison between pooled demographics of all officers (Column (7)) and the recruit demographics for each Exam period illustrates the changing nature of the Chicago Police Department. More recent recruits are less likely to be female than all panel officers (24.43% vs. the 2013 cohorts at 23.32%). While minorities make up roughly half of both groups, the composition of minorities has changed: Black officers make up about 22.75% of all panel officers while their share has been decreased by almost half from Exam 2006 (22.42%) to Exam 2010 (13.19%), to Exam 2013 (12.56%). The sharp decline in Black recruitment has been made up for by a surge in Hispanic recruitment (26.1% for all officers vs. 34.08% for Exam 2013). This pattern is generally representative of police departments across the country in the last 30 years ([141]). For a visualization of cohort compositions see Figure B.2.3.

Between Exam 2010 and Exam 2013, the reduction in start age requirement from 23 to 21 was associated with about a 2 year decline the average start age of recruits (Columns (1) and (5)). Figure B.2.4 displays the cumulative CDF of officer start ages in the three exam periods. Notably, in the Exam 2010 and 2006 cohorts, only 27% of recruits started before they were 27, while in the Exam 2013 cohorts 49% did.

The top panel of Table 2.2 displays summary statistics of main sample cohort compositions for all officers in Column (1), and Columns (2) and (3) divide these officers by whether their cohort had high ($\geq$ 50%) or low (< 50%) minority share (50% minority is the median). Low-minority cohorts are on average 43.8% minority compared with high-minority cohorts at 54.51%, with similar standard deviations– over all cohorts, one standard deviation of cohort share minority is 0.06. Comparisons of other demographic compositions show little differences in observables between low and high minority cohorts. The differences in cohort share female and mean start age are neither statistically significant nor economically large.

Policing Outcomes

Arrests are a common metric when studying individual officer and departmental performance and, in the light of concerns about over-policing — excessive and detrimental interactions between law enforcement and civilians — arrests are the main metric I use to measure officer enforcement activity.[16] To distinguish between the seriousness of arrests, I divide them based on crime: serious arrests, which I define as arrests for official FBI index crimes, as well as additional forms of homicide, fraud, domestic violence, sexual assault, and simple assault and battery; and low-level crimes are all other arrests—e.g., warrant, traffic, or drug crimes.[17] I also classify arrests based on arrestee race/ethnicity (white, Black, Hispanic, or other). Using Cook County court data, I determine if the arrest is associated with a guilty finding, and I interpret this as a measure of arrest

[16] Arrest counts (and the clearance of crimes) are a common metric of police activity. See [156], [194], [154], [195], [196], Blanes i [197], [161], [157], [145], and [198].

[17] Index crimes are offenses on which the FBI collects data and tracks and publishes annually in the Uniform Crime Report (UCR). The eight index crimes are four violent and four property offenses: (violent) aggravated assault, robbery, murder, rape, (property) burglary, larceny, motor vehicle theft, arson. For non-index crimes I classify as ‘serious’, domestic violence is determined by whether the description indicates domestic battery or assault, and a few additional sexual assaults were classified based on whether the description indicates criminal sexual assault. Simple assaults and battery include crimes such as attempts at assault, child abuse, and threats of violence. I classify multiple types of deceptive practices as fraud. As a robustness check in Section 2.5, I redo the main analysis using the FBI index and non-index crime distinctions for serious and low-level crimes, respectively, and find similar results.

quality.[18] Opportunities for officers to make arrests depend on the crime rates where and when they work, which influence the quantity, quality, and kind of arrests.[19]

The bottom panel of Table 2.2 displays arrests per shift, violent index crime rates, and observations in the daily panel data for all main sample recruits as full officers in Column (1), and Columns (2) and (3) divide these officers by whether their cohort had high ($\geq$ 50%) or low (< 50%) minority share.[20] Note that all differences in arrests and violent crime between Columns (2) and (3) are statistically significant. The vast majority of arrests are of Black civilians (81.8%), with Hispanic arrests being far less common at 13.2%. Recruits in high-minority cohorts make fewer arrests per shift than those in low-minority cohorts, driven by a difference in arrests of Blacks (0.1378 vs. 0.1481). About 70.31% of arrests are for low-level crimes. Recruits in low-minority cohorts make slightly more guilty arrests relative to those in high-minority cohorts, with 30.64% and 28.9% guilty, respectively. Recruits in low-minority cohorts work, on average, in slightly lower crime districts relative to recruits in high-minority cohort, yet both groups work in Chicago's most dangerous areas.[21] While this table documents differences between new officers in terms of arrest quantity and quality, as well as working environment based on cohort diversity, whether cohort diversity is actually changing officer enforcement behavior requires more detailed analysis.

[18] I define an arrest to be 'guilty' if the central booking number (CBN) is associated with any guilty finding; I consider an arrest not guilty if the CBN is associated with no guilty findings and at least one not guilty finding. If a CBN is associated with no guilty findings and no not guilty findings, and it has any dismissed cases, then I consider it dismissed. If a CBN does not appear in the court data, I classify the case as dropped. I group not guilty, dismissed, and dropped cases together and label them as 'non-guilty'. If a CBN is not classified as guilty, not guilty, or dismissed, but it is in the court data, then it only has incomplete/open cases, so it is classified as neither guilty nor non-guilty. A single CBN may have multiple charges or cases associated with it, and I use the method discussed above to provide a single outcome of an arrest which is conservative as only one guilty verdict on any charge is sufficient for an arrest to be 'guilty'.

[19] For example, lower crime may mean the marginal arrest is less likely to be high quality if officers value making arrests.

[20] Index violent crimes are murder, rape, robbery, and aggravated assault.

[21] A monthly violent crime rate of about 15 per 10,000 population is the 75th percentile of monthly violent index crime rates in Chicago.

2.3 Empirical Strategy

2.3.1 Peer Effects Framework

The aim of this paper is to estimate the long-run effect of peer diversity on officer behavior. The identification strategy for this paper borrows heavily from the education literature on long-run peer effects, leveraging the random assignment of officers (students) to academy cohorts (classrooms). As a first step, I adapt the regression specification from the long-run peer effects in education literature ([199], [171]) by regressing outcomes on the characteristics of randomly assigned peers. Specifically, I estimate:

$$\overline{Arrest}_i^k = \alpha_{p(i)}^k + \pi_1^k \overline{X}_{c(i)} + \pi_2^k X_i + v_i^k \tag{2.1}$$

where $\overline{Arrest}_i^k$ is the average number of arrests type k (e.g., Black low-level guilty arrests) per shift made by officer i randomly assigned to cohort $c(i)$ in exam period $p(i)$. Variable $\alpha_{p(i)}^k$ is a fixed effect for the exam $p(i)$ which officer i took (as did all other officers in i's cohort $c(i)$). X_i contains the demographic characteristics (e.g., race, start age) for officer i. $\overline{X}_{c(i)} = \frac{\sum_{j \neq i}^{n_c} X_j}{n_c - 1}$, contains the leave-out mean of the demographic characteristics of members of officer i's cohort c.[22]

The random assignment of lottery numbers within an exam pool allows cohort composition, $\overline{X}_{c(i)}$ to be uncorrelated with unobserved characteristics about the officer, v_i^k, permitting consistent estimation of the peer effect of cohort diversity, π_1^k (see Appendix B.1.2 for tests of random assignment). More formally: $\mathbb{E}[v_i^k | \overline{X}_{c(i)}, \alpha_{p(i)}^k] = 0 \; \forall \, i$.[23] However, the mechanism by which cohort composition influences future arrests is not specified. One part of π_1^k is the effect of cohort diversity on an individual officer's behavior, opinions, beliefs, and prejudices. Yet, as cohort diversity influences officer assignments and future peers (discussed more in Appendix B.1.4 and Appendix B.1.5), and assignments influence arrest possibilities, the other part of π_1^k is the assignment effect

[22] For computing $\overline{X}_{c(i)}$, I include all recruits beginning in the cohort c excluding i.

[23] Given that cohort composition is randomly determined and $\overline{X}_{c(i)}$ excludes the officer i, cohort composition excluding officer i is independent of officer i's observable characteristics, X_i. So, leaving out X_i should not impact estimates of π_1^k.

of diversity. So, though the assignment effect proves to be minor, π_1^k is a causal estimate of the effect of cohort composition on an officer's future arrests of type k *within* the assignment system of the Chicago Police Department.[24]

2.3.2 Officer Heterogeneity

Using a raw arrest metric, e.g., arrests per shift, as the outcome of interest is straightforward, but it suffers multiple issues. First, higher shares of minorities in cohorts cause officers to make fewer arrests of Blacks during their careers. This is in part due to how cohort diversity influences new officer assignments and how officers choose to bid for assignments (see Appendix B.1.4). Second, different cohorts do not start at the academy at the same time, and their opportunities for arrests will be influenced by departmental demand, non-linear changes in crime rates, and other factors, making arrests alone a highly noisy outcome.

For the effect of peers to be externally valid and relevant for police departments with different priorities and assignment policies, understanding how peers influence officer's individual *type* is necessary. By officer type, I mean their individual propensity to make an arrest, a measure of their enforcement activity regardless of their working environment, or their individual contribution to the quantity or quality of arrests they make. Controlling for working environment also alleviates concerns about differential assignments and crime rates over time.

Identifying the effect of peers on officer types follows two steps. First, I recover a measure of an officer's 'type', i.e. their propensity to make arrests net of high-dimensional daily assignment fixed effects and other factors. Second, I regress these arrest propensities on cohort composition to estimate the long-run effect of peer diversity on individual officer behavior. This allows for data reduction and exploration of heterogeneity, and it permits flexible specifications in the first stage.[25]

[24] In this setting, I cannot distinguish between endogenous and exogenous effects of peers ([200]), which means I cannot disentangle the effect of officers being affected by minority peers due to their behavior or their characteristics. I assume there are no correlated effects (e.g., instructor effects)– given the large amount of courses recruits are taught during the academy, it is highly unlikely that a cohort with 40% minority composition would receive different institutional environments or instructors than a cohort with 50% minority composition starting a month later.

[25] [145] uses an analogous method by first recovering officer fixed effects for making arrests following 911 calls, then regressing these officer fixed effects on officer characteristics, and [201] use a similar two-step procedure for studying the determinants of returns to education.

I first recover an estimate for all officers' (including those outside the full sample of cohort officers) propensities to make arrest of type k, θ_i^k, using a first stage regression.[26] I estimate a linear fixed effect regression model:

$$Arrest_{it}^k = \theta_i^k + \gamma_{brsw_t}^k + \beta^k V_{it} + \epsilon_{it}^k \tag{2.2}$$

where $Arrest_{it}^k$ is the number of arrests of type k officer i made during their on-duty time on date t.[27] The data is sufficiently rich such that I can control for a large set of assignment and environment characteristics with highly specific fixed effects, $\gamma_{brsw_t}^k$, which interacts officer i's assigned district and truncated beat code (b), their role (r), their shift number (s), and the year, month, and day of the week (w_t).[28] The data has over 7.8 million officer-shift observations on about 14,000 officers and contains approximately 580,000 assignment fixed effects (γ_{brsw_t}).[29] V_{it} controls for second-degree polynomials of officer i's tenure. All random shocks to an officer's arrest participation during their working period are contained in ϵ_{it}^k.

I assume that conditional on a polynomial of officer tenure and officer assignment fixed effects: 1) current and future shocks to arrest counts are orthogonal to past observables; 2) shocks to arrest counts are not serially correlated across shifts; and 3) since the number of daily shifts I observe for each officer grows quickly, the officer fixed effects are consistently estimated (the mean number

[26]More formally, $k \in \begin{pmatrix} All \\ Serious \\ low\text{-}level \end{pmatrix} \times \begin{pmatrix} All \\ Guilty \\ Non\ Guilty \end{pmatrix} \times \begin{pmatrix} Minority \\ Black \\ NonBlack \\ Hispanic \\ White \end{pmatrix}$.

[27]Estimation was performed using the R package 'lfe' (Gaure (2013a)), which implements the algorithm introduced in Gaure (2013b) that is designed for estimating linear models with multiple overlapping high-dimensional fixed effects (e.g. officers and who move across shifts or workers who move across firms). Notably, this package also allows for standard errors of the fixed effects to be recovered which is used in Appendix B.1.6.

[28]Formally, b is the numeric beat code with the last numerical digit removed, and r is the exact role designated by the full beat code. For example, beat code "2533" has a role of 'beat officer' and the beat is truncated to "253" which indicates the sector they work in (a group of contiguous geographic beats, and beats are on average less than 1 square mile). Beat code "2463A" has a role of tactical team C officer, as does beat code "2463C", and both have the same truncated beat as "246" (which does not map to a geographic sector), so their $brsw_t$'s are identical if they also work in the same watch, day of week, month, and year.

[29]While the assignment fixed effects are highly granular, there is still significant variation within and across assignments. There are on average 8 officers per value of γ_{brsw_t}, and on average officers work in 9 different b's and 5 different r's, for example. Thus, there is sufficient variation in assignments across and within officers to identify officer fixed effects.

of observations for an officer in the panel is 558). I interpret the recovered $\hat{\theta}_i^k$ as an estimate of the individual officer's propensity for enforcement of type k, which I recover for all officers in the daily assignment panel between 2010 and 2018.

With this first stage regression, I control for significant temporal, geographic, demographic, and income variation in where each officer is working, as well as the within-day heterogeneity, officer exposure to different types of civilians and local crime rates, and the effects of officer tenure.[30] Another strength of this design is that I am able to leverage data on all officers, not just those in the sample cohorts. So, I am using all the variation across officer shifts without having to drop observations from officers whose cohorts I do not observe. This means I have sufficient observations within highly granular assignments to use high-dimensional fixed effects (i.e. $\gamma_{brsw_t}^k$) and allow for interactions between assignment characteristics. Recovering a single metric (per arrest type) for each officer also avoids weighting issues as new cohorts have fewer observations in the panel data than older ones.

Using only the fixed effects of officers in my sample cohorts, I replace $\hat{\theta}_i^k$ as the dependent variable in equation (2.1):

$$\hat{\theta}_i^k = \alpha_{p(i)}^k + \pi_1^k \overline{X}_{c(i)} + \pi_2^k X_i + v_i^k \tag{2.3}$$

Now, π_1^k can be interpreted as the peer effect on an officer's propensity to make arrests of type k.[31] As before, the composition of one's cohort is independent of one's own pre-existing characteristics, but now the outcome is the result of extensively controlling for working environment such that $\hat{\theta}_i^k$ is officer i's individual contribution to make arrests of type k regardless of when or where they work. The minor effect of assignment crowding out due to cohort diversity is removed from this measure, and the exogeneity assumption ($\mathbb{E}[v_i^k | \overline{X}_{c(i)}, \alpha_{cp}^k] = 0 \forall i$) holds, making π_1^k the causal effect of cohort diversity on officer enforcement propensity. $\hat{\theta}_i^k$ is an estimated quantity (and thus has measurement error), and it is standing in for the ideal outcome variable in equation (2.3),

[30] The CPD's operational schedule reinforces the inability of officers to select shifts on specific days or civilian pools (see [144] for more detail).

[31] I assume a homogeneous treatment effect across officers.

θ_i^k, which is the officer's true but unobserved type. Following the teacher value added literature, I apply a Bayesian shrinkage procedure to the fixed effects to obtain more precise fixed effects and reduce measurement error (see Appendix B.1.6).

2.4 Results

2.4.1 Main Sample Results

The recovered distributions of main sample officer fixed effects indicate differences across race and exposure to diversity.[32] Figure 2.1 presents graphical evidence for heterogeneity in officer enforcement being related to officer race and cohort diversity. The distribution of officer arrest propensities (fixed effects) has a long right tail, as in [145]. Panel A displays the distributions of fixed effects for arresting Blacks for white and minority officers in the main sample. White officers tend to have higher fixed effects, i.e. a higher individual propensity to arrest Blacks, relative to minority (non-white) officers. This conforms with existing research on white officers policing Blacks more aggressively.[33] Panel B displays the distribution of fixed effects for white officers in the main sample split by cohort share minority. Clearly, white officers in cohorts with more minorities tend to have lower fixed effects—lower individual propensities to arrest Blacks.

Effect on Minority Arrests

For more detailed results, I turn to regression analysis. Table 2.3 displays the central results for the main sample. First, Column (1) displays the negative relationship between cohort share minority (CSM) and main sample officers' average arrests of minorities in their first 200 shifts (slightly over one year), estimated using equation (2.1). The coefficient on CSM, -0.188 (p = 0.057). This indicates that a 10pp increase in CSM (about 1.6 SDs) is associated with 4 fewer arrests of minorities over their first 200 shifts, equivalent to a 12% decline relative to the mean. Column (2) displays the effect of CSM on the officer's fixed effect (individual propensity after

[32] I solely discuss the fixed effects for officers in the main sample in this section, so "officer" or "recruit" both refer to officers in the main sample as sworn/full officers after their probationary period.

[33] See [143], [142], [145], and [144].

controlling extensively for assignment and temporal effects) to arrest minorities, estimated using equation (2.3). The coefficient is not statistically significant and smaller in magnitude (-0.138, $p = 0.11$), which is expected because of the negative bias in estimating equation (2.1) due to high CSM reducing the level of crime and minority population in an officer's working district (as shown in Appendix B.1.4).

The imprecision of Column (2) is due to significant underlying heterogeneity in the effect of minority peers on officers' propensities to arrest minorities. As shown in Columns (3) and (4), CSM has a large negative effect on propensity to arrest minorities for less serious (discretionary) crime but has a small positive effect on propensity to arrest minorities for serious (property and violent, non-discretionary) crimes. Based on Column (3), the coefficient on CSM (-0.186, $p < 0.05$) indicates that a 10pp increase in CSM decreases an officer's propensity to arrest minorities for low-level crimes equivalent to 1.86 fewer arrests over 100 shifts, which corresponds to a decrease 0.24 standard deviations and a 11% decline relative to the mean. Column (4), on the other hand, indicates that a 10pp increase in CSM increases an officer's propensity to arrest minorities for serious crimes by 0.48 more arrests over 100 shifts which is equivalent to 0.19 standard deviations. Columns (5) and (6) replicate Columns (3) and (4) with full controls for officer race, gender, start age, and cohort size; the results are almost identical with coefficients and standard errors increasing slightly. The robustness of the coefficients to these controls is further support for random assignment of officers to cohorts in the main sample.

The coefficients for officer race, gender, and start age, in Column (5) indicates that, for arrests of minorities for low-level crimes, relative to white officers, Black officers make 3 fewer arrests per 100 shifts while Hispanic officers make 1.6 fewer; male officers make 2.2 more arrests than female officers per 100 shifts, and officers who start the academy at one year older make 0.2 fewer arrests per 100 shifts. These differences by officer characteristics decrease in magnitude and statistical significance for serious arrests (Column (6)), as expected given that officers have more discretion over making low-level arrests as opposed to serious ones. These results are also in line with previous research about differences in aggressive policing by officer race, and the results are quite similar to those of [144].[34] Table B.2.1 provides more detailed results on the association between demographics and propensities for aggressive policing.

Columns (7) and (8) repeat the analysis in Columns (5) and (6) with CSM broken down into cohort share Black (CSB) and cohort share non-Black minority (CSN), which includes (mainly) Hispanics, Asians/Pacific Islanders, and Native Americans. Black peers have a significantly larger effect relative to non-Black minority peers on low-level arrests of minorities (-0.314, $p < 0.05$, vs. -0.193, $p < 0.05$) and serious arrests of minorities (0.097, $p < 0.01$, vs. 0.043, $p < 0.05$). The effect of CSB is about 2 times larger than the effect of CSN for both types of arrests.

Disaggregating Minority Arrests

Thus far, we have considered arrests of minorities as a single group which comprised just over 90% of arrests; however, 77.43% of minorities arrested are Black– the rest are almost all Hispanic (21.6%). Columns (9)-(12) break down arrests of minorities into arrests of Blacks ((9)-(10)) and arrests of non-Black minorities ((11)-(12)). Comparing Columns (9)-(10) and (11)-(12) shows that the effects on minority arrests are driven by arrests of Blacks, which maintains statistically and economically significant coefficients. A 10pp increase in CSB (CSN), equivalent to 2.1 (1.3) SDs, decreases officer propensity to make low-level arrests of Blacks by -0.0256, $p < 0.05$ (-0.0153, $p < 0.05$). This is equivalent to a 25.6% (15.3%) decrease relative to the mean. By contrast, the effects on non-Black minority arrests are directionally consistently but are small (for serious) and not statistically significant (for low-level). This is consistent with Table B.2.1, which shows that race, gender, and age have much stronger influences on Black arrests relative to non-Black arrests.

I focus on Black arrests as the main outcome of interest in the following sections because they drive the minority arrest results and are the vast majority of arrests in Chicago (>80%), though minority peers decrease officer propensities to arrest all civilians for low-level crimes.[35] Further-

[34] See [202], [160], [143], [145], and [142].

[35] Table B.2.2 shows that minority peers' decrease low-level arrests of civilians of all racial groups (Black, non-Black minority (mostly Hispanic), and white civilians), though point estimates are less precise for non-Black groups. The effects are largest and most precisely estimated for Black arrestees, which is consistent with the fact that Black civilians make up the vast majority of arrests.

more, I shrink officer fixed effects using a Bayesian shrinkage procedure described in Appendix B.1.6 for more precise results and use shrunken fixed effects in the following analyses.

Columns (13)-(14) replicate Columns (9)-(10), but using shrunken fixed effects. The effect of Black (non-Black minority) peers on low-level arrests shrinks by 22.27% (23.53%) and on serious arrests shrinks by 65% (78.79%), and standard errors decline. The fact that serious arrest fixed effects shrank more substantially than low-level ones is in line with the fact that they are much less frequent.

These final results indicate that a 10pp increase in CSB (CSN) decreases arrests of Blacks for low-level crimes by -1.99, $p < 0.05$ (-1.17,), over 100 shifts, equivalent to a -0.32 (-0.19) SD decrease. For serious arrests of Blacks, a 10pp increase in CSB (CSN) increases arrests of Blacks for serious crimes by 0.28, $p < 0.05$, (0.07, $p > 0.1$ over 100 shifts, equivalent to a 0.18 (0.05) SD increase.

Dynamic Effects

It is important to consider whether these effects are short-term or if they persist into an officer's career. While multiple years of observations are not available for full sample officers, the main sample officers are observable for a minimum of 3 years into their careers as full officers (4 years after they exit the academy). To explore the dynamic effects of peer composition, I re-estimate equation (2.2) with individual fixed effects for officers for each 180-day period– meaning an officer during the first 180 days in their career as a full officer is considered a different individual as that same officer in their next 180 day period. Figure B.2.5 displays the coefficients of the main specification (Column (13) in Table 2.3) estimated separately for each 180-day fixed effect for main sample officers. The figure indicates that the effect of non-Black minority peers slightly attenuates overtime, while the effect of Black peers remains stable. While 3 years is not an officer's full career by any means, the results indicate that the strongest effect (that of Black peers) is persistent long after the academy classes dissipate.

2.4.2 Full Sample Results

In this section, I expand the data used to the full sample– including the Exam 2006 and Exam 2013 cohorts. While adding additional observations is useful, the data come with limitations discussed in Section 2.2.2. Table 2.4 displays the results for the full sample, with the outcomes being shrunken officer propensities to arrest Blacks for low-level crimes (odd columns) and serious crimes (even columns), using equation (2.3) and including exam fixed effects.

Columns (1)-(2) display the results for Exam 2010 and Exam 2006 cohorts– the results do not change significantly relative to those in Columns (13) and (14) in Table 2.3. Columns (3)-(4) estimate the effects for the full sample (including Exam 2013). With the inclusion of the full sample, the effects of cohort share Black ($p < 0.01$) and non-Black minorities ($p > 0.1$) both decline and become less precise. Recall that the age requirement policy changed for the Exam 2013 cohorts, with the minimum start age being lowered from 23 to 21. This led to a significant compositional change between Exam 2010 (and Exam 2006) and Exam 2013 cohorts: there was a resulting shift in age distribution and the inclusion of a significantly higher portion of officers below the age of 27, as shown in Figure B.2.4 and discussed in Section 2.2.3.

Collectively, this raises the question of whether start age or minority status is driving the results. To disentangle the effect of age and minority status, I divide start ages into three groups: young (< 27), mid ($27-32$), and old (> 32), then I compute leave-out-means of each age group interacted with minority status (grouping Blacks, Hispanics, Asians, and Native Americans together as in Columns (1)-(6) in Table 2.3). Then I estimate equation (2.3) replacing cohort shares with these age and minority interaction groups. Controls include the share of mid and old white recruits, making the reference the share of young white peers. The results are displayed in Column (5)-(8), and they indicate that the interaction between age and minority status is a key factor with respect to the peer effects of diversity.

Based on the results in Column (5), the effect of minority peers is increasing by their age: young minority peers only slightly decrease arrests of Blacks for low-level crimes (-0.061, $p > 0.1$), mid-aged minority peers have a larger but noisy effect (-0.128, $p > 0.1$), and old minorities have an economically and statistically significant effect (-0.189, $p < 0.01$). Column (6) indicates that the effect of minority peers on serious arrests is also driven by older minorities (0.012, $p > 0.1$). This pattern in age groups is consistent with older officers policing Blacks less aggressively (see Table B.2.1). Columns (7) and (8) replicate (5) and (6) but excluding the Exam 2013 cohorts to ensure the shift in age distribution is not driving the results, and the patterns and results for the effects of minority peers are very similar for low-level arrests.

2.4.3 Effect on Officer Arrest Quality

A unique feature of my data is that I can observe the outcomes of arrests in court, which enables me to measure the quality of arrests. To study the effect of peer diversity on arrest quality, I classify high-quality arrests as those which result in a guilty outcome in court and low-quality arrests as those which result in a non-guilty outcome in court.[36] Much like arrest quantity, officers are heterogeneous in their arrest quality, and officers that have high propensities to make guilty (high-quality) arrests do not necessarily have high propensities to make non-guilty (low-quality) arrests. I estimate equation (2.3) on propensities to make non-guilty and guilty arrests separately. By comparing the effect of peer diversity on the propensity to make high-quality arrests with its effect on the propensity to make low-quality arrests, I can infer an effect of peer diversity on arrest quality separate from the influence of assignments, as crimes committed in some locations and times may be easier to arrest and prosecute than others. In Table 2.5, Columns (1)-(2) and (5)-(6) display the results for Black low-level arrests and Columns (3)-(4) and (7)-(8) display the results for Black serious arrests, with Columns (1)-(4) being main sample results and Columns (5)-(8) being full sample results, and guilty arrests in even columns and non-guilty arrests in odd

[36] As previously described, a guilty outcome means the arrest was associated with any guilty finding, and a non-guilty outcome means the arrest was not associated with any Cook County court case, resulted in only dropped charges, or had charges only found not guilty in court. Cases which had no final disposition or closed date in the data set are considered incomplete/open and are neither guilty nor non-guilty. Police officers do influence the initial charges against the arrestee and provide evidence and testimony to prosecutors and defendants, though their time in front of a jury or judge is limited particularly given the frequency of plea deals. Furthermore, while officer observables and behavior influence credibility in the eyes of judges, prosecutors, and juries, which alters their ability to make guilty arrests, an officer's cohort diversity is not observable to them and far removed. Additionally, officers may have reputations for being credible or not for prosecutors, meaning whether charges are filed may be a function of officer 'quality' as well. Lastly, a low-quality arrest does not necessarily mean the arrestee was innocent of any crime.

columns.

For the main sample, Black and non-Black minority peers have a small effect on officers making guilty low-level arrests, but they have large negative effects on officers' propensities to make non-guilty low-level arrests. To put the effects in context, the guilty / non-guilty rate for low-level Black arrests for main sample officers is 0.25, and similarly the ratio of the respective standard deviations in shrunken main sample officer fixed effects is 0.21. So, a 10pp increase in CSB decreases non-guilty low-level arrests by -0.0165 ($p < 0.05$) and guilty low-level arrests by -0.001 ($p > 0.1$), which corresponds to a guilty/non-guilty ratio decrease of 0.06. This indicates that CSB has a strong positive effect on arrest quality for low-level arrests because non-guilty arrests decline. Similarly, CSN has a small negative effect on non-guilty low-level arrests (-0.01,$p < 0.05$) and virtually no effect on guilty arrests.

For serious arrests of Blacks, only Black peers have any economically or statistically significant effect, with the coefficient on CSB for guilty serious arrests being 0.005, $p < 0.01$ and a noisy positive effect on non-guilty serious arrests, -0.002, $p > 0.1$. Given that the guilty/non-guilty ratio for serious arrests of Blacks is 0.33, this indicates Black (and non-Black minority, based on noisy coefficients) peers have a very small but positive effect on arrest quality for serious crimes.

This pattern of results is consistent in the full sample as well. Focusing on the interaction of age and minority status, Columns (5)-(8) display these results. Column (5) and (6) indicate that minority and mid- and old-white peers have a significant negative effect on non-guilty low-level arrests of Blacks and effects on guilty low-level arrests of Blacks more than order of magnitude smaller–indicating an increase in arrest quality. As in the previous section, the strongest negative effects are due to the share of older minority peers. For serious arrests, the effects are all statistically and economically insignificant indicating no significant change in arrest quality or quantity.

Low-quality arrests, however, are heterogeneous, as an arrest can result in a non-guilty outcome if 1. it results in a finding of not guilty (usually a judge ruling), 2. the case is dropped by the prosecutor or dismissed by the judge, or 3. it is missing from the court data.[37] In order to determine

[37] An arrest may be missing either because the officer later decided not to charge the individual (e.g. 'drunk tank' arrests or protesters), the individual was immediately sent to a diversion program, or due to matching error.

which of these is driving the low-quality results, I decompose low-quality, low-level arrests of Black civilians and redo my analysis on each sub-type. The results are displayed in Table B.2.3. The results show that dismissed/dropped cases and not guilty cases are driving the results, with no effect on missing cases, for both the main and full samples. This indicates that reduction is driven by cases with, for example, insufficient evidence or credible testimony by the officer, rather than matching error or the officer choosing to detain but not charge the individual with a sufficiently serious crime. This is consistent with officers being more discerning in their low-level arrests and thereby improving their average arrest quality.

2.5 Robustness

In this section, I discuss a variety of additional analyses that include alternative samples and specifications to test the robustness of the main results for both the main and full samples. Table 2.6 presents the main sample robustness tests and Table 2.7 presents those for the full sample, with each main sample test in Columns (1)-(7) being replicated for the full sample in the analogous column. Columns (8)-(11) in Table 2.6 present additional tests that are relevant or feasible for the main sample only. Due to the computational intensity of computing fixed effect standard errors, the fixed effects will be unshrunken except for Columns (7) and (11). Finally, I test the robustness of my results to common issues in peer effect studies.

2.5.1 Discrete Outcomes in First Stage

Arrests in a shift are count data, and the distribution of their frequency, as expected, fits a Poisson distribution. As such, I re-estimate the first stage (equation (2.2)) using a Poisson regression.[38] This model is potentially more reflective of the true data generating process, and environment and individual officer fixed effects likely contribute to arrests in a non-linear fashion. However, unlike the linear model, the estimates are not directly interpretable and fewer individual fixed effects can

[38] Specifically,

$$\mathbb{E}[Arrest_{it}^k|\theta_i^k, \gamma_{brsw_t}^k, V_{it}] = exp(\theta_i^k + \gamma_{brsw_t}^k + \beta^k V_{it}) \tag{2.4}$$

be recovered.[39] I use the recovered fixed effects in my second stage, and, as shown in Column (1) of Tables 2.6 and 2.7, the results are qualitatively similar to those of the main results, though not directly comparable due to the non-linearity of the model.

While the Poisson first stage is designed to more closely match the distribution in the data, a second concern may be that the results are driven by the skewed nature of the arrest data: most shifts have no arrests at all while very few have many. To test this, I re-estimate equation (2.2) as a linear probability model (LPM) with the dependent variable being if any arrest of type k was made by officer i during their shift. The results are displayed in Column (2) of Tables 2.6 and 2.7, and the estimates are very similar to the main results. These tests indicate that the results were not driven by either the reliance on a linear model in the first stage nor the skewed distribution of arrests per shift.

2.5.2 Alternative Samples and Controls

As multiple officers can be listed on a single arrest, this means some arrests are double counted in my analysis. This may be an issue if cohort share minority influences assignments in which only single officer arrests generally occur. To check the robustness of my results against this issue, I re-estimate the first stage only counting arrests for the first arresting officer. Column (3) in Table 2.6 displays the results of equation (2.3) on these recovered fixed effects, and they are similar to the main results but slightly smaller, as expected.

In the main results, the only control included is a polynomial of officer tenure as other factors, such as crime rates or shift duration may be endogenous to the officer's activity during their shift. As a robustness check, I re-estimate officer fixed effects with polynomials of local crime and watch duration during the officer's shift, and I use these officer fixed effects to use as the outcome in equation (2.3); I find the results are qualitatively similar (see Columns (4)).

While the assignment fixed effects control for significant amounts of heterogeneity in assign-

[39]The estimation is performed using the R package 'fixest' and an algorithm used to efficiently estimate fixed effects in maximum likelihood models ([203]). As the data are not overly dispersed, a negative binomial regression is not necessary.

ment, they do not control for the type of assignment an officer has, e.g. in car, on foot, or on bike patrol (though this is often determined by their beat and shift). To ensure that differential assignment types are not contaminating my results, I re-estimate equation (2.2) using the 91.9% of panel observations where an officer is in a car. The results are displayed in Columns (5) and are highly similar to the main results.

We used serious and low-level crime types in order to distinguish between types of arrests; however, we want to ensure that the results are not driven by this classification. I re-categorize the arrests based on the FBI index (serious: murder, rape, robbery, aggravated assault, burglary, theft, motor vehicle theft, and arson) and non-index (low-level: all others) classification and exclude warrant arrests as the exact crime type is not known. Fixed effects for Black index and non-index arrests are obtained by re-estimating equation (2.2). The results in Columns (6) are consistent with the main results.

Finally, with additional information about officers, I am able to add more controls about both officer and cohort compositions in order to test the robustness of the main results. These additional controls include officer level and cohort shares of Spanish speaking, female, high education (bachelors or above), and military-experienced officers as well as cohort mean start age. Columns (7) illustrates that the results are not significantly impacted.

2.5.3 Main Sample Only Tests

As previously discussed, the main sample cohorts have multiple advantages which also allow for more detailed exploration of effects, including additional robustness checks that are infeasible with full sample cohorts. First, the potential for non-Black officers to be sent to different assignments due to higher shares of Blacks in their cohorts may negatively bias my results (increasing magnitude) as high cohort diversity may lead to low fixed effects solely due to assignments (though the high dimensional working environment controls attempt to solve this issue). As a robustness test, I study the subset of new officers exposed to very high crime areas (above 75th percentile in violent index crime). As shown in Column (8) in Table 2.6, the 'high crime' new officers display similar results as the whole sample.[40]

Second, there may be a concern that new officers are not only assigned based on race and the racial composition of their cohort, but also based on their preferences and unobservables with respect to policing as well. I repeat my analysis on main sample officers during their probationary periods—which alleviates this issue as they have no actual policing experience upon entering this period– and I include controls for field training evaluator characteristics. Column (9) in Table 2.6 display these results, with effects qualitatively similar to the main results.[41]

Next, I repeat my analysis using a restricted sample similar to that of [144], which studied differences in policing outcomes between officers of different races and gender. Their sample differs in two main ways: first, it is restricted to 2012-2015 and excludes watch 4; second, it the assignment fixed effects used are more granular, controlling for the interaction between year-month, day of the week, shift, and exact beat code ('MDSB'), where as I interact assignment role with a truncated beat code.[42] I re-estimate the first stage on the restricted sample and control for assignment more granularly with MDSBs. Column (10) in Table 2.6 present the results which are similar to those of the main sample.

Lastly, while eligible applicants are permitted to enter the academy when their lottery number is drawn (and passing further tests), it is not required. A potential recruit may choose not to join the academy for a variety of reasons if their number is called. While all the estimates in this paper are conditional on uptake (i.e., only relevant for people who end up becoming officers), the composition of cohorts may be influenced by selection into the academy due to different start dates (though Appendix B.1.2 provides evidence against significant selection).[43] I test this by estimating

[40]This analysis is feasible in the main sample but not full sample as there is significant heterogeneity across exams in local crime rates and initial assignments, making it infeasible to decide on a sensible "high crime" cut off across 8 years of cohorts.

[41]Probationary field training officer data is only available fully for the main sample officer cohorts.

[42]The main analysis has other minor differences with their approach. I do not filter out officers for having additional information codes (e.g., indicating injury, training, union business, etc.) during their shifts as this information is not available for my full sample of assignment data. Naturally, this re-estimation is not feasible for Exam 2013 officers, which all start as full officers after this data ends in 2015.

[43]Accessing lists of individuals who did not enter the academy is not possible due to data privacy restrictions. However, for policy relevance, the effect of peer composition on the population of individuals who end up becoming police officers is likely more important than the effect of peer composition on the population on the margin of becoming police officers.

equation (2.3) on the subset of cohorts which started within 5 months of the initial main sample cohort, i.e. starting between July and December of 2012. Column (11) in Table 2.6 displays these results. The results for low-level arrests are qualitatively similar to those in the main results and the effects on serious arrests are insignificant. For further evidence, Figure B.2.6 displays how the coefficients of interest change as more cohorts are added.[44]

2.5.4 Measurement Error, Exclusion Groups, and Inference

Finally, there are three common issues with peer effect studies, particularly in settings that utilize similar assignment mechanisms as this paper. First, as discussed in [165], building off of [204], and in [205], peer effects can be over-estimated due to measurement error. In order to determine if measurement error in cohort composition is driving the magnitudes of my results, I follow [171] by adding measurement error to cohort racial compositions. Figure B.2.7 displays the results of adding increasing amounts of measurement error to the main specifications for the main sample and full sample. For the main sample (Panel A), increasing measurement error attenuates the main results, while in Panel B increasing measurement error modestly attenuates the main results. The Panel B results, which intersect age and minority status, are less attenuated by adding measurement error to race because no error is added to peer age, which also has an effect. Overall, this suggests that measurement error is unlikely to be driving the results.

Second, [165] also shows that there is a negative mechanical correlation between an individual and their peer composition, and this mechanical relationship may be driving the results. For the main specifications, I have used leave-out-means for peer compositions, so to determine if the mechanical correlation is driving the results I distinguish between the individuals being influenced from the peers who are doing the influencing ([206], [207], [171]). Table B.2.4 presents the results when excluding the peer group in question from the regression. Columns (1) and (2) present the effect of minority peers on white officers (excluding minority officers), Column (3) presents

[44] This is not feasible for the full sample because while about 40% of recruits in the main sample start within the first 5 months, it takes significantly longer for a sizable portion of Exam 2013 recruits to enter the academy, and there are only 5 cohorts in the Exam 2006 data set spread out across 2 years.

the effect of older peers on young and mid-aged officers (excluding older officers), and Column (5) presents the effect of female peers on male officers (excluding female officers); Columns (4) and (6) add interactions with officer race being white (see Section 2.6). The results are generally consistent with the results in Sections 2.4 and 2.6 (though generally noisier due to smaller samples), indicating that the results are robust to excluding the effecting officers.

Lastly, traditional inference techniques do not necessarily apply to many (quasi-)experimental designs, particularly peer effects studies where inter-group variation is a result of finite-sample bias. Recent peer effect studies use randomization inference to construct p-values for estimates ([208], [209], [174]), consistent with the guidance in [210]. I construct p-values using a randomization inference method which provides a distribution of estimates under the null hypothesis that there is no effect of peer composition on outcomes (see Appendix B.1.7 for more details). I construct 1,000 placebo cohorts with recruits randomly re-assigned at the exam level, then I re-compute cohort compositions and re-run the main regressions, and two-sided p-values are computed by ranking the coefficient in the main results within the distribution of placebo coefficients (in absolute value). As shown in Figure 2.2, the randomization inference p-values are generally smaller than those in the main results.

2.6 Mechanisms

There are multiple potential mechanisms underlying these results: positive interracial socialization, peer preferences, and negative interracial interactions. Overall, the evidence presented in this section is most consistent with the positive interracial socialization and the influence of peer preferences. I begin by elaborating on these two mechanisms, then I present evidence for them and discuss their implications. Finally, I discuss alternative explanations, such as peer race being correlated with other influential characteristics or instructors being influenced by the recruits they teach. To do so, I leverage the fact that additional peer characteristics, such as education and gender, are effectively randomly assigned by the lottery number system.

2.6.1 Primary Mechanisms

First, positive interracial socialization operates through contact between whites and minorities. During the academy, whites and minorities may become friends, causing a reduction in racial bias or prejudice among whites ([173], [174]). Evidence consistent with positive interracial socialization includes the effect of minority peers on whites being larger than the effect of minority peers on minorities.

Second, officers may adopt the preferences of those around them due to shifts in culture, social identity, or personality ([150], [152], [153]), meaning each recruit's preferences are influenced by the composition of their peers' preferences. For example, having more peers who prefer aggressive policing will cause an officer to police more aggressively in the future. While I can only observe officer behavior, I use arrest propensities as a proxy for officer preferences. Evidence consistent with the preference effect includes all officers being influenced by minority peers and the peer effect of a group being larger if that group, on average, polices less aggressively (e.g., the effect of female peers should be greater than male peers, as women tend to police less aggressively).

To determine the mechanisms behind the results, I first re-estimate equation (2.3) with additional terms interacting an officer being white with the variables of interest (minority cohort shares). Table 2.8 displays the results in Column (1) for the main sample and (2)-(6) for the full sample. I focus solely on low-level arrests of Black civilians because the main results for serious arrests were less precise and not economically significant. The net effect on whites can be calculated as the base coefficient (e.g., Cohort Share Black) added to the interaction term (e.g., White x Cohort Share Black), while the net effect on minorities is just the base coefficient (e.g., Cohort Share Black). Figure B.2.8 displays net effects on low-level and serious arrests visually for easier interpretation. For the main sample (Column (1)), the effects of Black and non-Black minority peers on white officers' propensities to arrest Blacks for low-level crimes are economically larger. There are multiple reasons why non-Black minority peers may reduce white officers' aggressive policing of Blacks: having more non-Black minorities (fewer whites) may facilitate Black-white contact and friendship; or contact with Hispanics, for example, may reduce bias against Blacks through the secondary transfer effect, documented in social psychology ([148]). Overall, the point estimates are consistent with both interracial socialization (whites being affected the most) and the influence of peer preferences (minorities are influenced as well); however, they are relatively imprecise, likely due to the small sample size.

For more precise estimates, I use the full sample for the remaining analyses. In Column (2), I include interactions for officers being white on their cohort shares of young, mid, and old minorities and mid/old whites. The effect of older minority peers on minorities is the largest (-0.166, $p < 0.01$), and the effect size decreases as the age group decreases. As older officers tend to police less aggressively (see Table B.2.1), this is consistent with the effect of peer preferences. By contrast, white officers' propensity to arrest Blacks for low-level crimes are larger and negative for all minority age groups, consistent with positive interracial socialization.

I further explore these mechanisms in remaining columns in Table 2.8. First, it is important to determine more precisely the relationship between age and minority race. In the main sample, Black peers had the strongest effect, and given that Black recruits tend to be older than non-Black recruits, we want to disentangle the race and age effects to see if the age group results are actually due to Black peers. In Column (3), I regroup minorities into Black and non-Black minorities and interact them with age groups young/mid and old (i.e., cohort shares for each combination of $\{\mathit{Black}, \mathit{Non\text{-}Black\ Minority}\} \times \{\leq 32, > 32\}$). The results in Column (3) indicate that although Black peers of all ages decrease arrest propensities, the largest and most precisely estimated effects are of older Black peers. The amplifying effect of age is also true for non-Black minority peers as well. So, while older minorities drive the effect of minorities on arrest propensities, older Black peers have an effect almost twice the size of older non-Black minority peers.

In Column (4), I add interactions with white officers, and the net effects are displayed in Figure B.2.8 for easier interpretation. As in Column (3), older Black and non-Black minority peers have the largest effects on minority officers relative to younger minority peers, with Black peers having a larger effect than non-Black minority peers. The effect of Black and non-Black minority peers on whites is larger than their effect on minorities, particularly for the effect of younger Black and non-

Black minority peers. These results are further evidence for interracial socialization influencing whites, and peer preferences causing shifts in all officers' future behavior on the job. Notably, in the full sample, even mid-age and older white peers have a significant effect on all officers, but no additional effect on whites, consistent with the fact that older white officers prefer less aggressive policing relative to younger whites (see Table B.2.1) but have no additional bias reduction effect on whites.

Gender is also correlated with minority status and age, so it may be that female minority peers are actually driving the effects. In Column (5), I add an additional control for cohort shares of white female and minority female recruits, and Column (6) adds the white officer interaction term. While the effects of minority and white peers by age group are similar to those of the Table 2.4, the additional coefficients indicate that female peers tend reduce arrest propensities, with white female peers having a small and not statistically significant effect (-0.049, $p > 0.1$) and minority female peers having an economically and statistically significant effect (-0.14, $p < 0.05$). This effect is in line with the fact that female officers tend to police Blacks less aggressively relative to male officers within race (see Table B.2.1 and [144]). Adding the race-white interaction term in Column (6) has little effect on the main coefficients, but it reduces the effect of female peers on minority officers. The coefficients suggest that female minority peers have a statistically and economically significant effect on all officers with a very small additional effect on white officers. However, white female peers have a small and very imprecise effect on minority officers and a relatively large effect on white officers. This pattern is consistent with inter-gender socialization being more common within race, and thus white females (who are less aggressive than white males) influencing their white male peers. While the interactions terms are imprecise, they are consistent with peers with propensities for less aggressive policing influencing all officers to police less aggressively.

There are three main implications. First, the evidence for interracial socialization reducing aggressive policing by whites highlights peer effects as one way in which departmental diversity alters officer behavior. Second, while policies such as procedural justice training have not produced significant changes in officer behavior ([211]) and common-place diversity trainings can be ineffective or counterproductive ([212], [213]), these results indicate that having diverse peers is an effective way of reducing aggressive policing. Finally, the effects of peer preferences are quite large. This indicates that while demographic representation is important, hiring diverse officers who have preferences for more aggressive policing (e.g., young and male) may actually be detrimental, as their peers may police more aggressively as a result.

2.6.2 Alternative Explanations

Although the evidence has thus far been consistent with positive interracial socialization and peers' preferences influencing officer behavior, there are alternative explanations. First, racial or age group classification could be correlated with other characteristics such as education or military experience. Education, for example, has been shown to be an important moderating factor in peer effect of diversity, as [174] finds that students assigned to peer groups with more mid- and high-performing Black students had increased future interactions with Blacks. Second, instructors may be influenced by class composition ([169]), thus changing how officers learn to police. For example, instructors with more Black students may treat them more respectfully, leading to officers policing Blacks less aggressively on the job. If education or military experience or instructors, and not peer racial composition, are driving the results, then the policy implications would be significantly different. Naturally, the evidence thus far completely rules out negative interracial interactions leading to animus as a mechanism.

For education and military experience, I explore these potential explanations in Table B.2.5 focusing on low-level arrests of Blacks using the full sample. Column (1) of Table B.2.5 displays the results from including additional controls for cohort share white and minority with high-education (a Bachelors or above) before the academy, and Column (2) adds the white officer interaction to these controls. In both cases, the variables of interest are unaffected. Peer education of whites or minorities does not seem to have any effect nor does it strongly influence the variables of interest of Table B.2.5. This is expected if officers are influenced by peer preferences because officer education is not associated with lower propensities for aggressive policing (see Table B.2.1). The interaction term in Column (2) for cohort share high education minority and officer race white is small, positive, and not statistically significant, while the interaction with cohort share high education white is positive and statistically significant. This indicates education is not mediating interracial socialization (i.e., whites are not only influenced by high education minority peers).

The same exercise is repeated for military experience in Columns (3) and (4). White military peers have a negative effect (-0.19, $p < 0.01$) and minority military peers have a positive effect (0.161, $p = 0.069$) on arrest propensities, which causes the effect of non-military minority peers of all age groups to be more precisely estimated and significantly larger. However, given that 97% of full sample recruits have military experience, the size of the effect is largely overstated. By comparison, the interaction terms in Column (4) are very small. Overall, being in the military attenuates peer effects, as military minority peers have smaller negative effects and military white peers have larger negative effects, but these effects are small in reality due to high military shares. From this and the education results, we can conclude that neither peer education nor military experience are driving the results, and they do not confound the previous evidence for the interracial socialization and peer preference mechanisms.

Finally, we turn to the possibility that peer effects of diversity are operating by influencing instructors, i.e., instructors behave differently in the presence of more minorities thus changing how recruits learn to police. While this cannot be directly tested as instructor identities are not known, there are multiple reasons why it is unlikely. First, this would likely be a very subtle effect, particularly given instructors are generally veteran police officers or related experts in law enforcement, who may not be particularly sensitive to class composition. This effect would likely also operate in courses related to conduct or ethics. Yet this training is quite minimal in the CPD curriculum, consistent with surveys on police training across the U.S. ([214]), implying instructor effects would be even more subtle.[45]

Furthermore, we would expect that if class composition is influencing instructors it would be strongest in small classes. For example, an instructor may be more likely to notice a 10pp increase

[45] [214] finds that on average only about 1% of training hours were dedicated to ethics, cultural competency, communication, or procedural justice each.

in share Black in a class of 20 recruits relative to a class of 70. In Appendix B.1.8, I test the effect of small class shifts, and the results indicate that it is unlikely that the effects of small-group composition are driving the results, which does not support instructor effects. Notably, the inclusion of Field Training Officer characteristics in Column (9) of Table 2.6 did not reduce the effect of minority peers during the probationary period– which would have been evidence of instructor effects during the probationary period. Overall, instructor effects, if present, are likely subtle and not the primary mechanisms driving the results.

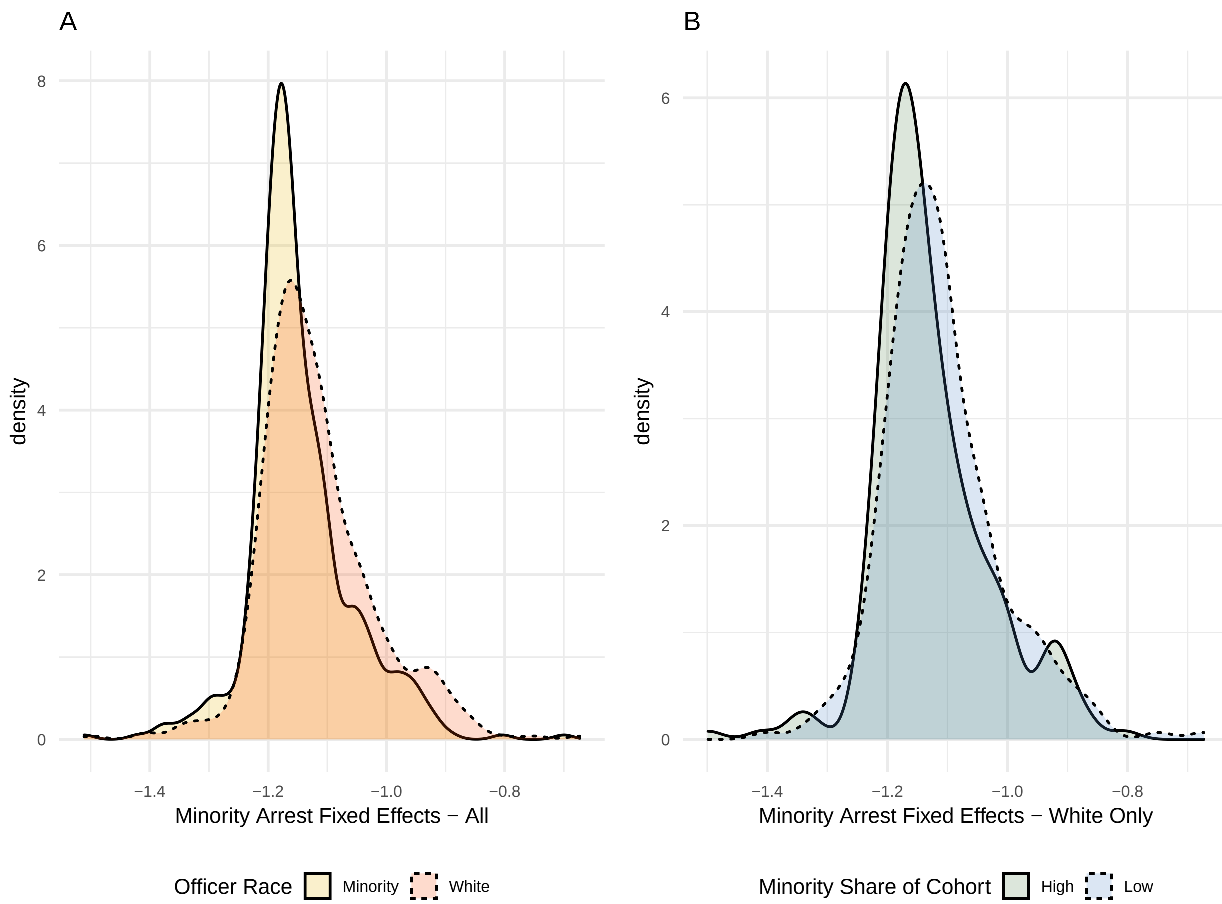
2.7 Conclusion

In this paper, I document the effect of minority peers in the police academy on police officers' future arrest quantity and quality. Higher shares of minority peers significantly reduce officers' future propensity to arrest minorities for low-level crimes. This effect is driven by a decline in low-quality (not resulting in a guilty finding) arrests of Blacks ($> 80\%$ of arrests), implying an increase in average arrest quality. Importantly, I find that minority peers have a small positive or null effect on arrests for serious crimes, implying officers are not reducing effort toward combating threats to public safety or property.

I find evidence consistent with two main mechanisms. First, white officers are most strongly influenced by minority peers of all age groups, consistent with interracial socialization reducing racial biases. Second, minority peers also reduce minority officers' propensity to police aggressively, though this is strongest for older and female minority peers, consistent with a shift in enforcement behavior due to academy peers' preferences. Notably, the bias reduction effect is smaller than the peer preference effect. I do not find evidence for instructor effects or negative interracial contact. Furthermore, the effects are present for at least 4 years after the academy ends.

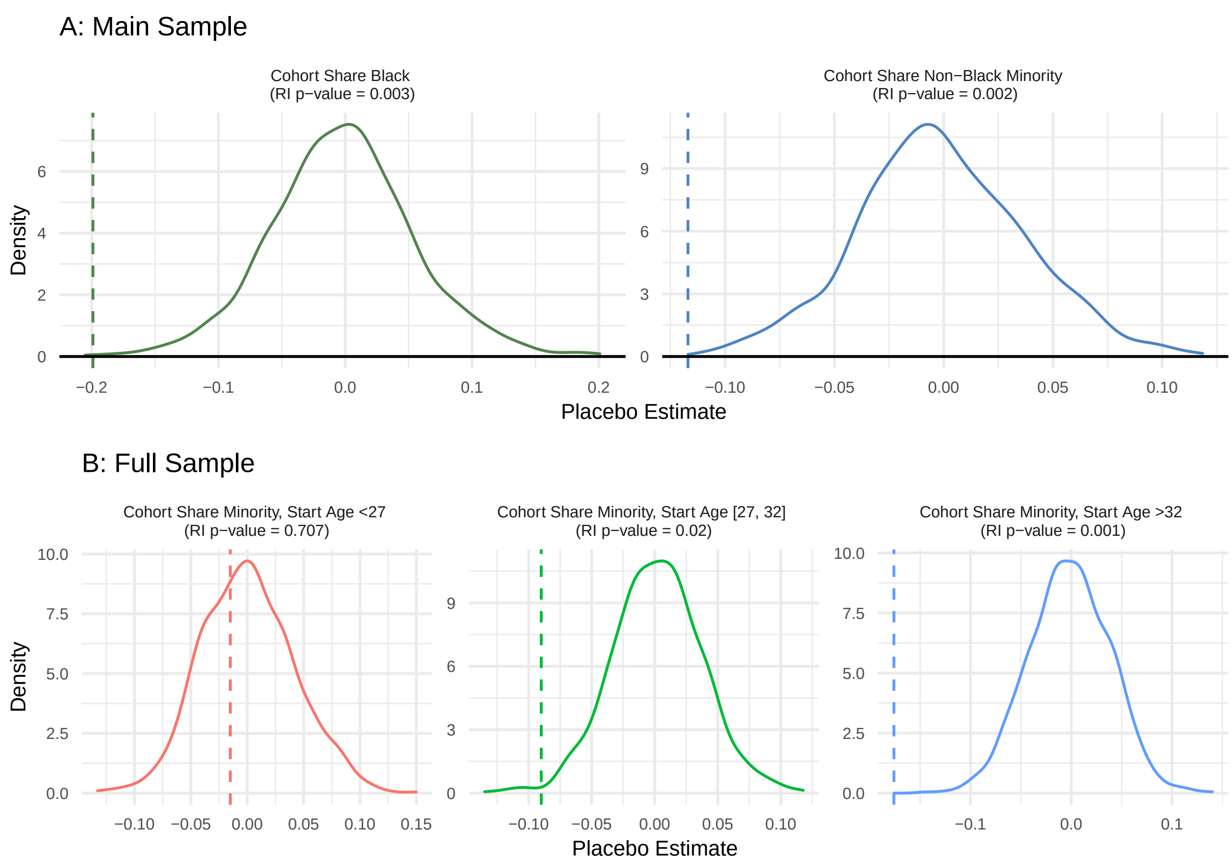
These results indicate that beyond minority status, race, gender, and age are all important factors which influence peer effects. In general, officers who police less aggressively themselves (e.g., Black, female, and older officers) cause larger shifts in their peers' propensities to police aggressively. If arrest propensities are indicative of officer preferences, then these results indicate that policy changes that result in more recruitment of minorities who prefer more aggressive policing may have self-defeating effects. For example, lowering minimum start ages in order to have more minority applicants may lead to increased recruitment of younger male minorities, which can nullify the effect of increased racial diversity due to the effect of increased peer preferences for aggressive policing. These results suggest that additional characteristics should be considered to improve departmental hiring policy, though this is ground for future research.

Overall, these results are generally consistent with the existing literature in economics and psychology on peer diversity in environments significantly different from a police academy. As the CPD and other major police departments all have sizable minority shares, the experiences of CPD officers are likely similar to those in other large cities. Furthermore, while Chicago is a single city, it provides an ideal environment because it is a sufficiently large department in a diverse city, which has a long-standing lottery system for the academy and maintains high quality data. Thus, the policy implications of these findings are far reaching and promising for improving policing. The inclusion of minority officers can result in persistent effects on their peers, reducing over-policing of low-level offenses while not reducing propensities to make arrests for more serious crimes, thereby potentially improving police-community relationships. Importantly, officers' arrest quality increases with peer diversity, so increasing departmental diversity through the recruitment of more minority officers can result in fewer wasted public resources, fewer individuals put under undue burdens, and fewer separated families.



Note: Figure displays the distributions of main sample officer (Exam 2010) fixed effects recovered from estimating equation (2) with arrests of minorities as the dependent variable. Panel A displays the distributions for white and minority officers separately, and it shows that white officers tend to have higher fixed effects for minority arrests. Panel B displays the distributions of white officers split by on whether they were in a high (at least 50%) or low (below 50%) minority cohort, and it shows that whites in high−minority cohorts tend to have lower fixed effects compared to whites in low−minority cohorts. Displayed fixed effects are generally negative due to the leave−out officer in the first stage having an arrest propensity higher than most main sample officers.

Figure 2.1: Distribution of Main Sample Officer Fixed Effects



Note: Figure visualizes the distribution of coefficients estimated using 1,000 placebo cohorts to conduct randomization inference, as discussed in Appendix A.7. Placebo cohorts are constructed by re−assigning recruits to cohorts within their exam period, ensuring the same number and size of cohorts. Coefficients are the effects of cohort shares Black and non−Black minorities in the main sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on shrunken officer fixed effects for arresting Blacks for low−level crimes. The dashed vertical lines correspond to the coefficient estimated in the main specification (actual cohorts), and the RI p−value denotes the p−value resulting from a two−tailed test which ranks the magnitude of the actual coefficient among the magnitudes of the 1,000 placebo coefficients. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sample).

Figure 2.2: Randomization Distribution of Coefficients (Reassigned Cohorts)

	Pooled	Cohorts	Pooled	Cohorts	Pooled	Cohorts	Pooled
	Main Sample (Exam 2010)		Exam 2006		Exam 2013		All Officers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2010	2010	2006	2006	2013	2013	
Male	0.807	0.8	0.808	0.769	0.767	0.778	0.756
Female	0.193	0.2	0.192	0.231	0.233	0.222	0.244
White	0.513	0.506	0.477	0.467	0.489	0.488	0.471
Minority	0.487	0.494	0.523	0.533	0.511	0.512	0.529
Black	0.132	0.13	0.224	0.229	0.126	0.121	0.227
Hispanic	0.312	0.325	0.263	0.273	0.341	0.35	0.261
Asian/Native American	0.0436	0.0396	0.0356	0.0319	0.0448	0.0407	0.041
Birth Year	1982.66	1982.7	1980.28	1980.34	1987.22	1987.15	1975.23
Start Age	30.07	30.16	29.6	29.54	28.18	28.04	28.99
Cohort Size	61.11	54.05	81.75	72	83.63	73.71	-
N	940	21	281	5	1115	17	11391

Note: Table compares the average characteristics of main sample (Exam 2010) officers (Columns (1)-(2)), Exam 2006 (Columns (3)-(4)), and Exam 2013 (Columns (5)-(6)) to all of the officers in the panel data (Column (7)). Column (1) contains the pooled average characteristics over all main sample recruits. Column (2) contains the average characteristics of the cohorts of the recruits in (1), including those recruits that do not appear in the main analysis due to attrition. Columns (3)-(6) replicate (1)-(2) for their respective samples. Column (7) contains the average characteristics of all officers in the daily assignment panel data.

Table 2.1: Summary Statistics by Sample

	All	High Minority Cohort	Low Minority Cohort	P-value
	(1)	(2)	(3)	(4)
Cohort Composition				
Share Minority	0.494 (0.064)	0.545 (0.031)	0.438 (0.036)	<0.01
Share Female	0.2 (0.039)	0.192 (0.039)	0.208 (0.04)	0.373
Mean Age	30.16 (0.542)	30.27 (0.565)	30.04 (0.515)	0.333
N. Cohorts	21	11	10	-
Officer Outcomes				
Total Arrests	0.175 (0.45)	0.172 (0.45)	0.178 (0.46)	<0.01
White Arrestees	0.008 (0.09)	0.008 (0.1)	0.008 (0.09)	0.01
Black Arrestees	0.143 (0.41)	0.138 (0.4)	0.148 (0.42)	<0.01
Hispanic Arrestees	0.023 (0.17)	0.025 (0.17)	0.021 (0.16)	<0.01
Total Serious Arrests	0.052 (0.25)	0.051 (0.24)	0.053 (0.25)	0.002
Total Low-Level Arrests	0.123 (0.38)	0.121 (0.38)	0.125 (0.39)	<0.01
Guilty Arrests	0.052 (0.31)	0.05 (0.3)	0.055 (0.32)	<0.01
Index Violent Crime Rate in District	15.4 (7.28)	15.08 (7.44)	15.72 (7.11)	<0.01
Obs	531597	262807	268790	-
Unique Officers	940	484	456	-

Note: Table presents the average cohort compositions at cohort level for main sample cohorts (top panel) and average number of arrests per shift, violent crime rate in average working district, and total observations for main sample recruits as full officers from 2013 to 2018 at the officer level (bottom panel). Columns (2) and (3) divide those recruits by whether or not they were in a cohort with a high (at least 50%) or low (less than 50%) share of minorities. Violent crime rate is determined by the district's average violent index crime rate per 10,000 population (2010 Census) in the month of an officer's assignment. P-values in Column (4) based on two-sided Student's t-Test comparing Columns (2) and (3). Standard deviations are reported in parentheses.

Table 2.2: Summary Statistics of Main Sample by Cohort Composition

	Avg. Arrests (First 200 Shifts)		Raw Arrest Propensity						Raw Arrest Propensity				Shrunken Arrest Propensity	
	Minority		Minority				Black				Non-Black Minority		Black	
	All	All	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Cohort Share Minority	−0.188*	−0.138	−0.186**	0.048*	−0.196**	0.044*								
	(0.099)	(0.086)	(0.091)	(0.025)	(0.097)	(0.024)								
Cohort Share Black							−0.314**	0.097***	−0.256**	0.080***	−0.058	0.017***	−0.199**	0.028**
							(0.148)	(0.028)	(0.117)	(0.024)	(0.036)	(0.006)	(0.098)	(0.013)
Cohort Share Non-Black Minority							−0.193**	0.043**	−0.153**	0.033*	−0.040	0.009***	−0.117*	0.007
							(0.098)	(0.021)	(0.074)	(0.018)	(0.026)	(0.003)	(0.061)	(0.009)
Black					−0.030***	−0.006**	−0.030***	−0.006*	−0.023***	−0.005*	−0.007***	−0.001	−0.021***	−0.004**
					(0.008)	(0.003)	(0.008)	(0.003)	(0.007)	(0.003)	(0.001)	(0.001)	(0.006)	(0.002)
Hispanic					−0.016***	−0.003*	−0.017***	−0.003*	−0.016***	−0.003**	−0.001	0.000	−0.014***	−0.003**
					(0.006)	(0.002)	(0.006)	(0.002)	(0.006)	(0.002)	(0.001)	(0.001)	(0.005)	(0.001)
Asian/Native American					−0.019	−0.007	−0.019	−0.007	−0.019**	−0.006*	0.001	−0.001	−0.017**	−0.004*
					(0.012)	(0.005)	(0.012)	(0.004)	(0.008)	(0.003)	(0.006)	(0.002)	(0.007)	(0.002)
Male					0.022***	0.007***	0.022***	0.007***	0.021***	0.007***	0.001	−0.000	0.019***	0.005***
					(0.006)	(0.003)	(0.006)	(0.003)	(0.005)	(0.002)	(0.002)	(0.001)	(0.005)	(0.002)
Start Age					−0.002***	−0.000	−0.002***	−0.000	−0.002***	−0.000	0.000	−0.000	−0.002***	−0.000
					(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cohort Size					0.000	−0.000*	0.000	−0.000*	0.000	−0.000*	0.000	−0.000	0.000	−0.000
					(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	0.249***	−1.064***	−1.389***	0.325***	−1.355***	0.340***	−1.341***	0.334***	−0.663***	0.269***	−0.678***	0.064***	−0.599***	0.190***
	(0.051)	(0.046)	(0.048)	(0.013)	(0.064)	(0.015)	(0.060)	(0.012)	(0.048)	(0.010)	(0.014)	(0.003)	(0.040)	(0.006)
R^2	0.008	0.008	0.019	0.012	0.077	0.048	0.081	0.055	0.082	0.053	0.023	0.017	0.074	0.043
Num. obs.	940	940	940	940	940	940	940	940	940	940	940	940	940	940

Note: Table displays the effect of cohort composition on main sample officer outcomes. The outcome in Column (1) is the average number of arrests of minorities (95 percent of all arrests) in the officer's first 200 shifts as a full (non-probationary) officer; effects estimated using equation (1). The outcomes of Columns (2)-(12) are individual officer fixed effects, estimated using equation (2), for all minority arrests, and minority, Black, and non-Black minority arrests for low-level and serious crimes; effects estimated using equation (3). Columns (13)-(14) replicate Columns (9)-(10) but use shrunken fixed effects, as described in Appendix A.6. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.3: Effect of Cohort Diversity on Arrest Propensity - Main Sample

	Shrunken Black Arrest Propensity							
	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	−0.211**	0.034***	−0.128*	0.026***				
	(0.096)	(0.012)	(0.069)	(0.006)				
Cohort Share Non-Black Minority	−0.115**	0.008	−0.041	0.005				
	(0.058)	(0.009)	(0.048)	(0.005)				
Cohort Share Minority, Start Age <27					−0.015	0.005	−0.061	−0.010
					(0.052)	(0.009)	(0.093)	(0.025)
Cohort Share Minority, Start Age [27, 32]					−0.090	0.005	−0.128	−0.004
					(0.071)	(0.007)	(0.097)	(0.015)
Cohort Share Minority, Start Age >32					−0.176***	0.021***	−0.189***	0.012
					(0.043)	(0.007)	(0.047)	(0.012)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Exams 2006, 2010	Exams 2006, 2010	Full	Full	Full	Full	Exams 2006, 2010	Exams 2006, 2010
R^2	0.232	0.115	0.422	0.154	0.428	0.154	0.235	0.113
Num. obs.	1221	1221	2336	2336	2336	2336	1221	1221

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes in even and odd columns, respectively. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3). Full controls refers to the additional controls in Column (13) of Table 3 with exam fixed effects and controls for cohort shares of whites who are start between 27 and 32 and above 32. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.4: Effect of Cohort Composition on Arrest Propensity - Full Sample

	Shrunken Black Arrest Propensity							
	Low-Level		Serious		Low-Level		Serious	
	Guilty	Non-Guilty	Guilty	Non-Guilty	Guilty	Non-Guilty	Guilty	Non-Guilty
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	−0.010	−0.165**	0.005***	−0.002				
	(0.013)	(0.073)	(0.002)	(0.008)				
Cohort Share Non-Black Minority	−0.003	−0.105**	0.000	−0.003				
	(0.007)	(0.046)	(0.002)	(0.004)				
Cohort Share Minority, Start Age <27					0.005	−0.026	0.000	0.000
					(0.005)	(0.043)	(0.002)	(0.004)
Cohort Share Minority, Start Age [27, 32]					−0.002	−0.091	−0.002	−0.002
					(0.006)	(0.058)	(0.002)	(0.004)
Cohort Share Minority, Start Age >32					−0.009**	−0.154***	−0.002	0.000
					(0.004)	(0.036)	(0.001)	(0.003)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Main	Main	Main	Main	Full	Full	Full	Full
R^2	0.068	0.074	0.031	0.036	0.107	0.496	0.033	0.032
Num. obs.	940	940	940	940	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to make high (guilty) and low (non-guilty) quality arrest of Blacks for low-level and serious crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.5: Effect of Cohort Composition on Arrest Quality Propensity

	Poisson	LPM	First Arresting PO	Crime and Watch Duration Controls	Car Only	FBI Index/Non-Index	Additional Controls	High Crime	Probation	Restricted Sample	First Cohorts
	Panel A - Black Low-Level Arrests										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cohort Share Black	-4.77^{***}	-0.21^{**}	-0.15^{**}	-0.24^{**}	-0.28^{**}	-0.22^{**}	-0.21^{**}	-0.33^{***}	-0.18^{***}	-0.24^{***}	-0.19^{**}
	(1.79)	(0.10)	(0.07)	(0.11)	(0.12)	(0.10)	(0.10)	(0.11)	(0.06)	(0.08)	(0.09)
Cohort Share Non-Black Minority	-2.90^{**}	-0.13^{**}	-0.09^{*}	-0.14^{**}	-0.18^{**}	-0.14^{**}	-0.19^{**}	-0.12^{*}	-0.11^{**}	-0.15^{***}	-0.25^{***}
	(1.19)	(0.06)	(0.04)	(0.07)	(0.08)	(0.06)	(0.08)	(0.06)	(0.04)	(0.05)	(0.03)
R^2	0.14	0.09	0.07	0.08	0.08	0.08	0.09	0.09	0.09	0.10	0.09
	Panel B - Black Serious Arrests										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Cohort Share Black	3.48^{***}	0.06^{***}	0.05^{***}	0.06^{***}	0.08^{***}	0.07^{***}	0.04^{**}	0.07^{*}	0.10^{***}	0.14^{***}	-0.00
	(0.85)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.02)
Cohort Share Non-Black Minority	1.87^{***}	0.02^{*}	0.02	0.02	0.03^{*}	0.04^{***}	0.00	0.04	0.03	0.03	-0.01
	(0.59)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Controls	Full	Full	Full	Full	Full	Full	Additional	Full	FTO Demos	Full	No Cohort Size
R^2	0.13	0.05	0.05	0.05	0.05	0.08	0.05	0.06	0.03	0.06	0.03
Num. obs.	890	940	940	940	940	940	940	509	940	908	385

Note: Table displays the effect of cohort diversity on officer propensity to arrest Blacks for low-level (Panel A) and serious crimes (Panel B) for main sample officers. The propensity is captured by officers' fixed effects using equation (2), Column (6) uses shrunken fixed effects. The parameter estimates are based on the specification in equation (3) (unless otherwise specified), with controls denoted as Full referring to the specification in Column (13) of Table 3, and additional or removed controls denoted in the table. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses.

Columns (1),(2): Results from estimating officer fixed effects using equation (4) and modifying equation (2) with the dependent variable (arrests of type k) being whether (1) or not (0) the officer made at least one arrest of type k during their shift. Column (3): The fixed effects used as dependent variables were recovered from estimating equation (2) with an arrest only contributing toward the dependent variable if the officer was the first arresting officer.

Column (4): The fixed effects used as dependent variables were recovered from estimating equation (2) including crime (violent, property, sex, drug, domestic, and other) and watch duration second degree polynomials in the estimation.

Column (5): The fixed effects from estimating equation (2) on the sample of shifts where officers were in cars (>85Column (6): Reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types. Column (7): Uses shrunken fixed effects (e.g., used in Columns (13)-(14) of Table 3) with additional controls for officer being in the military, speaking Spanish, having a Bachelors degree or higher and cohort shares of females, military, Spanish speakers, and Bachelor degrees or higher, and cohort mean start age. Column (8): Results use new officers from the main sample whose average district of assignment was above the 75th percentile of violent index crime in Chicago. Violent index crime rates are violent index crimes (murder, rape, robbery, aggravated assault) in a month, based on Chicago City Data Portal crime data, per 10,000 population, based on 2010 Census estimates.

Column (9): Results are from estimating equation (3) on fixed effects of new officers recovered during their probationary periods only, with additional controls for share of Field Training Officers (FTOs) that were Black, Hispanic, and other (non-white) race.

Column (10): Results from estimating the first stage (equation (2)) on a restricted sample of officer assignments, only including assignments in 2012 to 2015, with no additional information codes, and only regular watch assignments (the three main shifts), using assignment fixed effects as described in [144].

Column (11): The sample is the subset of the main sample of recruits who started within 5 months of the first 2012 cohort, which is 7 cohorts. Fixed effects are shrunken. Due to few cohorts, [215] standard errors are used clustered at the cohort level, and I do not control for cohort size.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.6: Alternate Samples and Specifications - Main Sample

	Poisson	LPM	First Arresting PO	Crime and Watch Duration Controls	Car Only	FBI Index/Non-Index	Additional Controls
				Panel A - Black Low-Level Arrests			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohort Share Minority, Start Age <27	−1.04	−0.03	−0.03	−0.04	−0.05	−0.03	−0.04
	(1.37)	(0.06)	(0.04)	(0.07)	(0.08)	(0.06)	(0.06)
Cohort Share Minority, Start Age [27, 32]	−2.84*	−0.10	−0.09	−0.12	−0.15	−0.09	−0.10
	(1.48)	(0.08)	(0.05)	(0.09)	(0.10)	(0.08)	(0.07)
Cohort Share Minority, Start Age >32	−5.43***	−0.21***	−0.14***	−0.23***	−0.29***	−0.21***	−0.23***
	(1.11)	(0.05)	(0.03)	(0.05)	(0.06)	(0.05)	(0.07)
R^2	0.73	0.53	0.43	0.50	0.58	0.49	0.44
				Panel B - Black Serious Arrests			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohort Share Minority, Start Age <27	1.30	0.03	0.02	0.03	0.03	0.03	0.01
	(0.97)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Cohort Share Minority, Start Age [27, 32]	1.27	0.04**	0.03*	0.03*	0.04**	0.04**	0.01
	(0.98)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Cohort Share Minority, Start Age >32	3.91***	0.07***	0.06***	0.07***	0.09***	0.09***	0.03***
	(0.83)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Controls	Full	Full	Full	Full	Full	Full	Additional
R^2	0.72	0.41	0.44	0.37	0.46	0.71	0.16
Num. obs.	2207	2336	2336	2336	2336	2336	2336

Note: Table displays the effect of cohort diversity on officer propensity to arrest Blacks for low-level (Panel A) and serious crimes (Panel B) for main sample officers. The propensity is captured by officers' fixed effects using equation (2), Column (6) uses shrunken fixed effects. The parameter estimates are based on the specification in equation (3) (unless otherwise specified), with controls denoted as Full referring to the specification in Column (13) of Table 3 and controls for cohort shares of whites who are start between 27 and 32 and above 32, and additional or removed controls denoted in the table. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses.

Columns (1),(2): Results from estimating officer fixed effects using equation (4) and modifying equation (2) with the dependent variable (arrests of type k) being whether (1) or not (0) the officer made at least one arrest of type k during their shift. Column (3): The fixed effects used as dependent variables were recovered from estimating equation (2) with an arrest only contributing toward the dependent variable if the officer was the first arresting officer.

Column (4): The fixed effects used as dependent variables were recovered from estimating equation (2) including crime (violent, property, sex, drug, domestic, and other) and watch duration second degree polynomials in the estimation.

Column (5): The fixed effects from estimating equation (2) on the sample of shifts where officers were in cars (>85Column (6): Reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types. Column (7): Uses shrunken fixed effects (e.g., used in Columns (13)-(14) of Table 3) with additional controls for officer being in the military, speaking Spanish, having a Bachelors degree or higher and cohort shares of females, military, Spanish speakers, and Bachelor degrees or higher.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.7: Alternate Samples and Specifications - Full Sample

	Shrunken Low-Level Black Arrest Propensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	−0.121					
	(0.119)					
Cohort Share Non-Black Minority	−0.043					
	(0.053)					
White x Cohort Share Black	−0.150					
	(0.110)					
White x Cohort Share Non-Black Minority	-0.144^{**}					
	(0.059)					
Cohort Share Minority, Start Age <27		0.006			−0.012	−0.013
		(0.052)			(0.051)	(0.051)
Cohort Share Minority, Start Age [27, 32]		−0.038			−0.087	−0.088
		(0.067)			(0.067)	(0.067)
Cohort Share Minority, Start Age >32		-0.166^{***}			-0.165^{***}	-0.166^{***}
		(0.050)			(0.041)	(0.040)
White x Cohort Share Minority, Start Age <27		−0.052				
		(0.038)				
White x Cohort Share Minority, Start Age [27, 32]		-0.113^{**}				
		(0.050)				
White x Cohort Share Minority, Start Age >32		−0.021				
		(0.053)				
Cohort Share Black, Start Age <=32			−0.091	−0.083		
			(0.120)	(0.110)		
Cohort Share Black, Start Age >32			-0.231^{**}	-0.207^{**}		
			(0.113)	(0.092)		
Cohort Share Non-Black Minority, Start Age <=32			−0.034	0.012		
			(0.050)	(0.043)		
Cohort Share Non-Black Minority, Start Age >32			-0.142^{***}	-0.116^{**}		
			(0.045)	(0.049)		
White x Cohort Share Black, Start Age <=32				−0.028		
				(0.087)		
White x Cohort Share Black, Start Age >32				−0.058		
				(0.131)		
White x Cohort Share Non-Black Minority, Start Age <=32				-0.102^{**}		
				(0.042)		
White x Cohort Share Non-Black Minority, Start Age >32				−0.053		
				(0.050)		
Cohort Share Minority and Female					-0.140^{**}	-0.114^{*}
					(0.064)	(0.061)
Cohort Share White and Female					−0.049	−0.001
					(0.093)	(0.100)
White x Cohort Share Minority, Female						−0.052
						(0.052)
White x Cohort Share White, Female						−0.094
						(0.081)
Cohort Share White, Start Age >=27		-0.101^{***}	−0.064	-0.082^{**}	-0.067^{*}	-0.067^{*}
		(0.034)	(0.046)	(0.040)	(0.039)	(0.039)
White x Cohort Share White, Start Age >=27		0.043		0.030		
		(0.027)		(0.034)		
Controls	Full	Full	Full	Full	Full	Full
Sample	Main	Full	Full	Full	Full	Full
R^2	0.079	0.430	0.428	0.430	0.433	0.434
Num. obs.	940	2336	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3) with the addition of terms for cohort shares interacted with an officer being white. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.8: Interaction of Cohort Diversity and Officer Race

Chapter 3: The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago

Bocar Ba and Roman Gabriel Rivera

3.1 Introduction

After each viral police killing, a familiar response cycle occurs: public outrage leads to calls for police reform and increased oversight and these proposals are countered with predictions of higher crime if officers are more intensely monitored. The predictions of increased crime stem from a “de-policing" theory ([216, 217] or the “Ferguson effect" ([218], which propose that increased oversight will drive officers to reduce effort, resulting in higher crime rates. Yet, empirical evidence for such effects generally relies on highly public scandals as instruments for increased oversight ([219, 220]).[1] As a result, studies which link oversight and higher crime using public scandals ignore the effect of a scandal on civilian behavior, and their estimates of the effect of oversight on crime are contaminated by the simultaneous effects of public outrage.

In this paper, we argue that the effect of public scandals on crime, misconduct, and other policing outcomes is due to both oversight and outrage effects. We use detailed data on misconduct, arrests, stops, and uses of force from the Chicago Police Department (CPD) in combination with an interrupted time series (ITS) design ([225]) to study the effects of oversight events and an outrage event on officer and civilian behavior. We follow [226, 227]'s method to bound our estimates without an external control group. To study the effects of oversight and outrage on crime, we apply synthetic control ([228]) and synthetic difference-in-differences methods ([229]) to national Uniform Crime Reporting (UCR) data. For each event, we measure public awareness using local

[1] See also [221, 222, 223, 224].

news coverage, and we use scraped discussions from a popular Chicago police forum and monthly union newsletters to measure officer awareness.

To guide our analysis, we outline a simple conceptual framework for distinguishing between the effects of oversight and outrage. We define an increase in oversight as a change that increases the effective cost of misconduct on officers– e.g., more transparency, higher likelihood of reporting or punishment, increased sanctions. Oversight directly decreases officer misconduct either by decreasing engagement with civilians to avoid error-based misconduct or by decreasing deliberate acts of malfeasance. If misconduct is driven by officers making errors when exerting effort to reduce crime ([220]), then an increase in oversight will cause officer effort and misconduct to decline and crime to rise. However, if misconduct can be deliberate and counterproductive, then the effect of oversight on effort and crime is unclear ex-ante. Outrage, on the other hand, influences both civilian and officer behavior, and only partly through increases in oversight. In times of public outrage, crime rises because public criticism and social unrest (e.g., protests) reduce officer morale and strain departmental resources and community distrust and social disorder increase. Reported misconduct may decline due to reduced deliberate-misconduct and officers reducing contact with civilians (and thus error-misconduct), or it may rise because civilians are highly incentivized to report misconduct. As we draw a clear distinction between oversight and outrage, we note that instruments for changes in oversight are valid if they influence officer behavior by increasing (or decreasing) the effective cost of misconduct, and they do not independently influence outcomes through other mechanisms.

With this framework, we provide empirical evidence that a public scandal influences both officer and civilian behavior. In late 2015, footage of the officer-involved-shooting of a Chicago teenager, Laquan McDonald, was released and caused a major public scandal in Chicago accompanied by massive public outrage: protests, political shifts, and a DOJ investigation. We find that this ‘outrage’ caused reductions in officer interactions with civilians (lower discretionary arrests, stops, uses of force) and increases in crime, particularly murder. Our estimates suggest that after the scandal, violent crime in Chicago rose 19% in the first 5 months, and the scandal caused 17 additional murders per month over the subsequent 15 months. These results are inline with the existing scandal-based literature. However, we find evidence that civilian behavior also changes: despite reduced police-civilian contacts, complaints remain flat– consistent with misconduct being driven down by reduced interactions and officers avoiding deliberate-misconduct but reporting being pushed up by the shift in civilian incentives to report.

Our main contribution, however, is identifying the effect of oversight, not accompanied by outrage, on policing activity, crime, and misconduct. We identify two events linked to unexpected court rulings that increased oversight (effectively increased the cost of misconduct) in Chicago but were largely unknown by the public. The first event, the ruling in *Gekas v. Williamson*, occurred in July of 2009, and the second event, the police union's response to the ruling in *Kalven v. Chicago*, occurred in July of 2014. Crucially, we show that while CPD officers were highly aware of and sensitive to these events, the public paid little to no attention at the time. Following each 'oversight' event, complaints against police officers declined– specifically complaints filed by civilians– by between 28% and 34% in the first 5 months. However, policing activity did not change significantly: arrests, stops, and use of force were largely unchanged. Importantly, crime was also unaffected. Collectively, these results suggest that oversight itself (unaccompanied by outrage) does not lead to increased crime or reduced officer effort.

Panel A of Figure 3.1 illustrates the main point of our paper. We report the number of allegations of misconduct filed against Chicago Police officers over time and contrast it with the number of violent crimes. The two first vertical lines correspond to events that impacted the level of officers' oversight, while the third event corresponds to the officer-involved-shooting which triggered public outrage. Overall, both outcomes decline over time, except after the last event, where we show that the level of misconduct allegation remains constant while the level of violent crime spiked. In contrast, following the first two events that only increased the level of oversight the level of complaints declined without changing the number of reported violent crimes.

Our results provide empirical evidence for recent theoretical work on misconduct. Building on the results of [230], we find evidence that alleged victims are strategic in reporting misconduct, and we find that civilian behavior toward the police is not exogenous during volatile periods– despite evidence of less aggressive enforcement and contacts with civilians, complaints remain flat. During periods of outrage, victims recognize that the police department is more likely to respond to police complaints and that they have the support from various groups (ACLU, DOJ, etc.), so are more willing to come forward.

This paper contributes to several strands of the law and economics literature. Since [231], researchers have found evidence suggesting crime responds to policing tactics and presence.[2] However, as argued in [243], there is no reason to assume that officers never engage in malfeasance while enforcing the law. Hence, officer misconduct could impact multiple dimensions of policing through suspensions, overly aggressive policing, racial profiling, or damaged community relations. Our paper contributes to this literature by measuring the impact of changes in oversight and outrage on relevant outcomes such as crime, arrests, and misconduct allegations.

Previous works using public scandals or protests have studied the effect of police illegitimacy on murder rates and the effect of racial uprisings on police killings ([244, 245]), but they have not focused on policing outcomes such as complaints or arrests. Other studies that estimate the effect of oversight on officer conduct have used scandals in quasi-experimental designs that generated increased media attention and judicial scrutiny; they find increasing oversight decreases police effort and increases crime rates ([219, 220]). We argue that one cannot interpret such estimates as the effect of oversight on policing because scandals induce simultaneous and significant changes in civilian incentives which impact crime rates, through mechanisms such as social disorder and legal cynicism.[3] This is informed by existing empirical evidence: [247] documents that high-profile cases of police violence result in citizens being less likely to report crimes, and this can increase crime opportunities independent of police effort; [248] finds that legal cynicism influences criminal behavior; [194] finds that officer damaged morale and unmet pay expectations decrease effort and increases crime; and [249] shows that police killings damage nearby students' educational

[2]See [232, 233, 234, 235] on staffing levels, [236, 237, 238, 239, 240, 234, 241, 242] on deployment and tactics, and [155, 161] on problem oriented policing.

[3]"Legal cynicism refers to a cultural orientation in which the law and the agents of its enforcement are viewed as illegitimate, unresponsive, and ill equipped to ensure public safety." ([246])

attainment and mental health.

We extend this literature in multiple ways. We find further evidence that scandals increase crime. However, we provide a conceptual framework that allows researchers to disentangle oversight from outrage. In effect, we propose that the scandal-based literature is identifying an outrage effect which should not be interpretted as the effect of oversight. Lastly, we find new evidence that, in absence of a public scandal, oversight resulting from court rulings and union memos can reduce misconduct without impacting crime or police effort. This also relates to a literature showing how officers are highly responsive to managerial directives ([194]). In particular, studies found that more internal monitoring leads to higher hit rates for stop-and-frisk in New York ([250]) and fewer violent crimes in New Orleans ([251]). Our results are also consistent with [252], which shows that a public attention matters in policing as a decrease in public focus on crime causes clearance rates to decline.

Lastly, this study advances the policing literature by focusing on the relationship between misconduct and policing outcomes. There is ample theoretical literature on the cost of enforcement against innocent individuals (e.g. false arrests or failures to provide service) by [156, 66, 253], which contributes to understanding how officers may engage in misconduct as they enforce the law ([243, 254]). However, with the exception of [255] and the literature on body cameras,[4] studies on the costs and benefits of proactive policing have generally focused on racial differences ([66, 265, 266, 267]).[5] As such, our main contribution is finding that oversight does not necessarily increase crime or reduce officer effort.

The rest of the paper is organized as follows. Section 3.2 introduces a conceptual framework which distinguishes between and makes predictions of the effects of oversight and outrage. Section 3.3 discusses the data. Section 3.4 provides background on the oversight and outrage events.

[4]See [256, 257, 258, 259, 260, 261, 262, 263, 264]. The studies on body worn cameras produce mixed results, with some evidence that they improve public trust and reduce the use of force, while others find no effect on use of force but increase assaults on officers.

[5]There is a large literature providing compelling empirical research designs to test for the presence of racial bias in motor vehicle stops and searches ([268, 269, 270]). However, those designs do “not allow for the possibility of false accusation by police or planting of evidence" (See footnote 5 in [268]). [271] provide legal and theoretical explanations of officer incentives to make searches, seizures, and false arrests.

Section 3.5 presents the empirical strategies. Section 3.6 presents the results, and Section 3.7 probes their robustness. Section 3.8 discusses the implications and concludes.

3.2 Conceptual Framework

In this section, we present a conceptual framework which incorporates police oversight and public outrage, and makes predictions of their effects on reported misconduct, crime, and officer effort. We draw distinctions between the conceptual framework presented in this paper and those of existing papers.

We define oversight as any system or event which increases the effective cost of misconduct for police officers. Similar to deterrence of crime, oversight is composed of the sanctions associated with being accused or found guilty of misconduct and the likelihood of one's misconduct being detected. Thus, oversight can come in many forms: larger punishments for misconduct, more thorough or strict investigations by oversight agents, increased civilian reporting of misconduct, etc.. Naturally, as oversight increases officers will try to decrease misconduct as the net benefit of misconduct has decreased. However, the fundamental question is how does this change in officer incentives influence their effort and crime.

Many commentators and researchers have envisioned misconduct as unintentional officer errors, as in [220], while others see misconduct as an outcome of confrontations with suspects which can result in genuine errors or accusations of an error by guilty suspects, as in [219]. For example, an officer can accidentally strike a suspect, arrive at a call late by accident, or arrest the wrong person. We call this error-misconduct (EM). Under this formulation, because error-misconduct is a byproduct of officer effort to deter crime and catch criminals, the only way for officers to reduce misconduct is to reduce effort. Thus, the effect of oversight on officer behavior and crime is clear ex-ante: increasing oversight leads officers to reduce misconduct by decreasing effort, and crime rises as a result.

We propose that misconduct is composed of error-misconduct and deliberate-misconduct (DM). As opposed to EM, DM is any form of misconduct that is separate from officer effort. Furthermore, DM is counterproductive as it involves officers allocating time to illegitimate activities that damage trust with civilians and are not linked to crime reduction– indeed, DM can include actions that are crimes themselves. DM encompasses not responding to 911 calls, failing to submit crime reports, illegal searches, excessive force on detained suspects, and using racial slurs. Under this formulation, it is not clear ex-ante how oversight influences crime or officer productivity: increased oversight will cause officers to decrease total misconduct, affecting both EM and DM; while decreasing EM pushes crime up, decreasing DM pushes crime down as officers reallocate time from misconduct to more productive activity. Thus, the net effect on crime depends on the magnitudes of and substitution between each effect, and its sign not clear ex-ante.

Next, we draw a distinction between oversight and what we term 'outrage'. The existing literature use major scandals as instruments for police oversight because in the face of public outrage the cost of misconduct increases ([219, 220]). However, we propose that public scandals are not valid instruments to study the effect of oversight on crime or officer behavior, because public outrage has a multitude of effects separate from oversight. First, public outrage leads to criticism and scapegoating of the police which damages morale; as [194] shows, morale is an important driver of officer effort and thus crime prevention. Second, outrage causes social disorder, most visibly protests, which require large police presences and strain departmental resources, which emboldens criminal elements. Indeed, the original formulation of the 'Ferguson effect' by Ferguson Police Chief Sam Dotson discusses resources being strained due to civil unrest and that the department would need significantly more officers to keep crime down.[6] Third, outrage is accompanied by decreased trust in the police among civilians, and civilians with worse views of police are less likely to follow the law ([248]). Lastly, in the face of a scandal, civilian cooperation with police and reporting of crimes falls ([247, 272]), which makes criminal opportunities more attractive.

The effect of outrage on crime is positive as the large public reaction (and its many consequences) push crime up. However, the effect of outrage on reported misconduct is not clear ex-ante: while officers reduce misconduct (as outrage also increases oversight), civilians simul-

[6] See here.

taneously increase their likelihood of reporting misconduct because the social benefit is higher in a period of outrage and because civilians are strategic reporters ([230]). Our framework predicts that following an outrage event, crime will rise, officer activity will decline, but reported misconduct may experience a muted effect in either direction. We also predict that the effect of oversight alone (without outrage) will reduce reported misconduct but may have little effect on crime and officer behavior due to the opposite effects of EM and DM on crime.

3.3 Data

In this section, we discuss the various data sets that we use to explore the effects of oversight and outrage. There are three main reasons why it is notably difficult to quantify the effect of changes in oversight. First, police oversight agencies are relatively rare in the U.S.; even if a department has an oversight agency, accessible misconduct data is rare. Beyond this, different agencies have different areas of authority, definitions of misconduct, classifications, etc., making comparisons between cities difficult. Second, actual misconduct is not possible to observe, meaning reported misconduct (much like reported crime) must be used as a proxy (despite the possibility changes in oversight regimes can change civilians' incentives to report misconduct). Third, oversight events, much like outrage events, are qualitative in nature.

Complaints We measure reported misconduct using a dataset of misconduct complaints obtained via Freedom of Information Act (FOIA) requests as part of a lawsuit, *Kalven v. the City of Chicago & the Chicago Police Department* (Kalven v. Chicago). The data contains the universe of misconduct allegations filed against CPD officers between 2001 and June 2016. This dataset was the first major collection of police misconduct records obtained and released to the public. The authors were involved in the collection, cleaning, and release of this data in collaboration with the Invisible Institute.[7] Unfortunately few other cities have followed and released comprehensive datasets on complaints. Due to the lack of standardization across oversight agencies in the US and poor record

[7] With the Invisible Institute, we also helped build the website CPDP and distributed the data to the public and other researchers. This background provides us with deep understanding of the data's strengths and limitations.

keeping, complaint data is nearly impossible to compare across jurisdictions. Overall, this makes the CPD complaint data from this period a uniquely rich data set on oversight and misconduct.

This data set is the final of multiple iterations obtained from the CPD in November of 2016. Later releases of CPD complaint data (containing cases after June of 2016) were subject to alterations by the replacement oversight agency in Chicago or the CPD, which we believe significantly damaged the integrity and internal consistency of the data. For example, some cases closed by the city oversight agency between 2007 and 2016 were reopened or the allegations were altered and complaint types were updated. Complaints were retroactively added to the data as well.[8] For this reason, we use the complaints data released in November of 2016 as it is internally consistent and contains records and cases unaltered by the new agency.

One of the main advantages of the data obtained through *Kalven v. Chicago* is that complaints are categorized by the type of alleged misconduct. We categorize misconduct complaints into three main categories (in order of severity): Serious (i.e., constitutional violations), Failure to Provide Service (FPS), and Other. As a single incident may involve multiple individual allegations, a complaint is identified by its highest level of severity and whether any officer was identified in the complaint. We include complaints with identified officers and known complainants unless otherwise specified.

Table 3.1 provides summary information on the 20 most common types of complaints (85% of all complaints), including their share of total complaints, likelihood of being filed by a fellow officer (not a civilian), and likelihood of the complaint being sustained. The table illustrates a few important facts relating to the 78,000 complaints (at the officer-allegation level) with incidents occurring between 2007 and May 2016. First, the most common complaint relates to failure to provide service, which tend to be filed by civilians seeking help from the police, for example

[8]For example, when comparing the public data from the Invisible Institute (complaints-complaints and complaints-accused from 2016 and 2018): CR#1015193 exists in the 2016 data and is not in the 2018 data; CR#1030541 has a complaint date 2 months later in the 2018 data compared with 2016 data, and is closed in the 2016 data with no closed data and no accused officers in the 2018 data; CR#265893 for UID#265893 (officer ID) has a final finding of sustained in the 2016 data but is reported not sustained in the 2018 data; in the 2016 data, CR#1002368 for UID#100040 was an off-duty domestic altercation (category 05K), while in the 2018 data the category and description have entirely changed.

the officer failed to file a report or did not respond to a call for service. Three of the top five most common complaints relate to search; and about 12% of all complaints are related to searches (either vehicle or person) without a warrant– both incidents which were unlikely to be errors on the part of the officer. Racial and ethnic complaints, such as verbal abuse and use of slurs, are the 17th most common type of complaint, followed by off-duty domestic violence, and the 20th most common is off-duty unnecessary use of force.

Notably, police officers only file a small share of complaints (about 10% overall), with some of the largest shares relating to conduct unbecoming of an officer on duty, submission of improper reports, and inventory procedures. The sustained rate for complaints is highly correlated with an officer filing the complaint: while only 1.5% of civilian filed complaints are sustained, almost 40% of police filed complaints are sustained. For almost all complaint types, over 90% have a known complainant– excluding complaints with missing types, which 71% only have complainants.

Crime One set of crime data is provided by the City of Chicago which reflects reported incidents that occurred in Chicago from 2001 to the present (at the time of writing). Each incident contains information about the crime, such as location, date, type, and whether or not an arrest was made. This data includes both index and non-index crimes, which we separate into drug crimes and other non-index crimes.[9]

In addition to the Chicago-specific data, we supplement this data with UCR data on index crimes across the US. This dataset is made usable by [273], who constructed a monthly dataset on a police department-level basis for index crimes (excluding arson). Because the UCR program does not systematically provide the number of reported rapes for Chicago, we do not include this outcome in our analysis. We use this data to compare Chicago index crimes with other cities as a control group.

For both the Chicago data and UCR data, the main outcome of interest is the number of reported crimes per 100,000 residents (i.e. crime rate). Because of the presence of zeros in the

[9]Index crimes include violent (murder, rape, aggravated assault/batter, robbery) and property (burglary, larceny theft, motor vehicle theft, and arson). Non-index and non-drug crimes include prostitution, misdemeanors, traffic crimes, etc..

data, especially for murder, we chose to analyze per 100,000 population levels rather than logs or percentage changes. See Appendix C.1 for more details.

Arrests, and Stops, and Use of Force To measure officer interactions with civilians and effort, we use data on arrests, street stops (beginning in 2010), and use of force. Arrests are a commonly used measurement of officer effort in economics ([194, 220]), and we add to this with stops and use of force as well. Unfortunately, the UCR data does not systematically provide the number of crimes cleared by arrest (much less stops or uses of force for any department) for the Chicago Police Department. Thus, similar to the case of complaints, we are limited in terms of our ability to find a suitable external control group for these outcomes.

Public Awareness To measure the public's awareness and attention dedicated to policing and misconduct in Chicago, we use a collection of 8,000 scraped *Chicago Tribune* articles relating to the Chicago Police Department. We classify these articles based on whether or not they mention topics such as crime, misconduct, police accountability, or oversight agencies (e.g., DOJ, COPA, IPRA) (see Figure 3.1).

Police Awareness To measure the CPD's awareness and response to events, we use scraped posts from Second City Cop (SCC) and monthly newsletter published by CPD patrol officer's union, the Fraternal Order of Police (FOP) Lodge 7. SCC is a (now offline) forum for CPD officers to anonymously discuss work, news, and drama within the department without a filter. SCC is useful in studying anonymous discussions of officers, as in other professional fields with similar forums ([274]).

The monthly union newsletter, on the other hand, provides insights into what top department and union members want to tell officers and give us an understanding of the messaging of 'the brass'. Separate from these events, the newsletter provides crucial information for officers such as relevant contract negotiations, policy changes, fundraisers for fallen officer families, pension news, etc..

3.4 Background on Events

In order to distinguish between oversight and outrage, we must identify relevant events for both cases. As evidenced by the scandal-based literature ([219, 220, 275]), major policing scandals are relatively common– and have become increasingly common in recent years. In this paper, we use a scandal (Event 3) as an 'outrage' event. However, our main contribution is rooted in identifying 'oversight' only events, which are far more elusive. These oversight events (Events 1 and 2) have the important feature that they increased police oversight while not causing other major shifts in public attitudes or civilian incentives. This section provides a timeline of the events we are studying backed by public documentation from the Chicago FOP newsletter and SCC posts.

3.4.1 Oversight: Events 1 and 2

Events 1 and 2 are related to court cases which influenced officer's perceived cost of misconduct. Event 1 consists of a quick sequence two court cases and a report from the CPD's union. On July 20, 2009, the ruling in *Gekas v. Williamson* indicated that "internal affairs files are a public record regardless of the outcome of the probe" ([276]). This unexpected ruling had "repercussions for long-running complaints about Chicago police brutality" ([276]).

Almost immediately, CPD officers took notice of this ruling. On July 24, 2009, in a post entitled "Danger Will Robinson! DANGER!!!" and beginning with "This has the potential for being bad. Very bad" (linking to an article about the ruling) was posted on SSC. The poster laments the dangers of this rulings for CPD officers, and concludes with "The Local and State FOP better get on the ball here. The National Lodge might not be a bad idea either". The comments, which counted over 70 in the following two days, expressed significant concern and frustration over the ruling. Figure 3.5 displays a screen shot of the post, and the entire post and 19 pages of the comments section is available at this link, however it contains profanity and some disturbing language. Clearly, the officers understood that this significantly reduced their ability to have their complaints be sealed, forgotten, and hidden from the public.

Indeed, a major component of this is that internal affairs files (such as complaints) were ruled to be public record *regardless of the outcome*. As shown in Table 3.1, almost all complaints by civilians are not sustained, meaning if only sustained or disciplined complaints had been made public only a small subset of officers and complaints would be vulnerable to exposure. Yet, including all internal affairs files in the public record meant that all officers with complaints would be vulnerable. For context, between 2007 and June of 2009, over 7,000 CPD officers had complaints filed against them, and between 2007 and June of 2016 it was almost 12,000 CPD officers had a complaint– and this does not even include the large number of complaints prior to 2007.

At the same time, the CPD was being sued for violating Diane Bond's constitutional rights when interacting with her in *Bond v. Uteras*. It was ruled on in November of 2009, and further opened the gates for future lawsuits on making misconduct files public. In the November 2009 union newsletter, the union came out with a clear message on misconduct investigations reminding officers about their rights to request to secure legal counsel from the FOP if they are involved in an allegation of misconduct. In addition, in the Second Vice President's Report in the newsletter (see Figure 3.6), the report warned officers of the severity and potential risks of being named in a complaint investigation. Importantly, this short sequence of events occurred long before the current debate on police reform.

Event 2 is the fallout resulting from the *Kalven v. Chicago* lawsuit (filed in November 2009). While the ruling occurred in March of 2014, the police union did not respond until July 28, 2014,[10] when the police union "filed an Emergency Motion requesting a Temporary Restraining Order" to challenge the release of some of the names in the lists of officers with misconduct records.[11] In the August 2014 newsletter, the front page article (the President's Report) discussed the ruling and its implications for officers (see Figure 3.7). It stated that "In compliance with the court decision, lists of past complaints against Chicago Police Officers will now be released for public review." With this, began a period of significant concern over the release of complaints (referred to as CRs) by both officers and higher ups. As stated on the front-page of the September newsletter, under

[10]Newsletters between March and June 2014 do not mention the *Kalven v. Chicago* ruling.
[11]https://www.illinoiscourts.gov/Resources/61e475d6-97b0-4eef-8d62-23766553a041/1121846.pdf

the header “CRs and FOIA", the president’s report states: “For the past several weeks now we have been consistently speaking to the Membership on the particulars concerning the Freedom of Information Act requests that seem to have taken on a life of their own", then after discussing how easy it is to file a FOIA request, it states: “We find this totally unacceptable and will work on how to best attack this oversight by the courts during the next legislative session in Springfield" (see Figure 3.8).

Both Events 1 and 2 increased officers’ cost of being accused of misconduct, and thus increased oversight. This cost was not due to tangible effects, such as a more stringent oversight body or higher penalties, rather they sharply increased the social and professional cost of misconduct, similar to [277]’s ‘extralegal’ cost of being caught for a crime. Complaints are almost never sustained or disciplined (see Table 3.1), and even when complaints are sustained the punishments are rarely significant or even enacted, so the tangible cost of misconduct is quite low for police.

However, as stated by the FOP, in October 2014, if complaint records were made public officers would suffer “public humiliation and loss of prestige in their employment".[12] Essentially, public complaint records meant that accusations of domestic violence, racial abuse, sexual harassment, and other activities would become public record and viewable by not only random civilians, but coworkers, superiors, friends, and family members. Furthermore, complaints damage officers’ professional abilities; for example, in court, defense attorneys can use lists of (unsustained) complaint allegation against an officer in order to question their credibility in the eyes of the judge and jury. Given that complaints carried only a very low tangible cost, this social and reputational cost was highly salient, as evidenced by not only the officer reaction but how aggressively the union fought the release of such records. Furthermore neither ruling had an immediate effect, rather they increased the likelihood of future releases of complaints, increasing the future cost of present (allegations of) misconduct.

To use these events as instruments for oversight, however, they must, like all instruments,

[12] See https://www.chicagotribune.com/news/ct-fop-police-lawsuit-met-20141028-story.html . This was the response following the city agreeing to satisfy the request for records (see https://news.ballotpedia.org/2020/06/30/illinois-supreme-court-rules-that-chicago-police-disciplinary-misconduct-records-cannot-be-destroyed/).

satisfy relevance and exogeneity. As just discussed, the events likely significantly increased the effective cost (reputational and social) of being accused of misconduct for officers, and thus are relevant. We also provide visual evidence of this change. As shown in Figure 3.1, both Events 1 and 2 preceded declines in the number of complaints received by police officers, further supporting the likelihood that officers changed their behavior in response to these events.

In order to satisfy exogeneity, however, we must ensure that these events did not change civilian incentives or trigger an outrage-like period. Indeed, both of these events were, at the time, largely unnoticed by the public. Figure 3.1 displays the share of articles in the Chicago Tribune by topic. There is virtually no change in discussion of the CPD around either of these events.

3.4.2 Event 3: Outrage

Finally, we contrast these two events with the public outrage of Event 3. Event 3 is the public release of the dash-cam footage of the murder of teenager Laquan McDonald by a CPD officer which occurred in November 2015 and immediately brought national attention ([278]) and scrutiny onto the department including coverage by national newspapers such as the New York Times (see Figure 3.9).[13]

The massive public response to the scandal is illustrated by the jump in the number of articles related to the CPD in Figure 3.1. Unlike the oversight events, the number of articles related to police accountability and police misconduct all substantially increased during the months following Event 3, leading to a level-increase in all such topics. Ultimately, the scandal led to a civil rights investigation by the Department of Justice (DOJ) and resulted in the implementation of multiple reforms to the Chicago Police Department and its oversight system by consent decree. [279] provides a detailed discussion of this incident and its implications for policing. Moreover, the number of articles mentioning the DOJ or the police union also increased sharply during this month. Indeed, the mayor at the time, Rahm Emanuel, over five years after the incident, was questioned about his handling of the scandal by Congress in 2021.[14] Figure 3.11 provides a discussion from

[13]The shooting occurred in October 2014 but drew no media attention until the footage was released.

[14]https://www.chicagotribune.com/politics/ct-rahm-emanuel-ambassador-senate-confirmation-hearing-20211020-

the president of the FOP related to the DOJ investigation in January 2016's newsletter.

Moreover, the public release of the complaints in November 2015 due to Events 1 and 2 coincided with Event 3. As a result, the FOP president wrote an article in the November 2015 newsletter (see Figure 3.10) discussing the complaint data and related Events 1 and 2 to the public release.

The swelling public outrage is clear in the coverage of the CPD by the *Chicago Tribune* (and other news outlets). Much like the other scandals in the existing literature, Event 3 is an ideal example of 'outrage', and we will use it in contrast with the oversight events. Given the significant difference in public scrutiny between Events 1 and 2 (which drew virtually no public attention) and Event 3 (which resulted in public outrage, protests, and political shifts), we argue that Events 1 and 2 could only have induced a change in CPD officer behavior, while public outrage toward police and focus on misconduct increased sharply following Event 3.

3.5 Empirical Strategy

We want to study the effect of oversight and outrage on a variety of outcomes: crime, complaints, and enforcement. As discussed in the data section, the complaint data and much of the enforcement data is uniquely detailed in Chicago which makes finding comparison cities nearly impossible. Crime data, however, is widely accessible which allows for alternative methods to be used.

As the events affect the same city and police department with a few years between them, we can interpret the effect of each event as cumulative. That is, the observed effect of Event 2 is in addition to the effect of Event 1, and Event 3 occurs in the context of both Events 1 and 2. Event 1 raised the possibility of complaint records becoming public in the future and increased the perceived severity of complaints for officers. Event 2, in the context of Event 1, increased the likelihood of complaint records becoming public and easily accessible (by FOIA requests) from a possibility to a near certainty. And Event 3 generated a public scandal in the context of successful litigation against the CPD for open complaint records (Events 1 and 2). In fact, the release of the video (resulting from

dozfx5efdvcyxppidyywl52gle-story.html

an open records request and lawsuit) was only possible due to the court decisions relating to Event 2, which itself was only possible because of the court decisions surrounding Event 1. Furthermore, the events are qualitative in nature so we cannot determine their relative salience, only that they were sufficiently salient to produce a response by officers and, in the case of Event 3, the public.

3.5.1 Chicago-Level Analysis

We begin our analysis with the Chicago-only data. We follow [225] by using an interrupted time-series design to estimate the short-run effect of a change in oversight and outrage. This method compares the average outcome τ months prior to the event e occurring in month $e(T)$ to the first $\tau - 1$ months following the event:

$$\theta_{e(T)} = \frac{1}{\tau - 1} \sum_{t=T+1}^{T+\tau} Y_t - \frac{1}{\tau} \sum_{t=T-\tau}^{T-1} Y_t \qquad (3.1)$$

We exclude the month of the event itself as they are partially treated. Furthermore, we residualize the outcome using month fixed effects in order to control for seasonality. We assume that the outcome (residualized) would have been constant in the months just before and after each event if the event had not occurred. While this is a strong assumption, for most events and outcomes, after removing seasonality, it is consistent with the visual evidence (see discussion in Section 3.6). In our main specification, $\tau = 6$, though we test the robustness of the results using larger and smaller τ in Section 3.7. So, the ITS design recovers the short-run (within 6 months) effect of each event on the level of the outcome using a comparison of post-event Chicago to pre-event Chicago. While the ITS design estimates a change in level, we do not believe the mechanisms behind each event occurred instantaneously.

To assess the level of uncertainty of our estimates, we recompute the test statistic for placebo months available in our sampling period. We define placebo windows as those with none of the events. For $\tau = 6$, there are 70 placebo months. We report the 5th and 95th percentile values from the permutation tests and the p-values associated with each test. The p-value for the null hypothesis that the event has no effect such that the change we observe occurred at random is

$1/(70+3)$ (including the actual event dates).

3.5.2 Chicago vs Rest of the U.S

The main challenge for understanding police oversight in our context is that the oversight change took effect all at once in Chicago. Thus, other factors might have influenced the outcomes of interest before the change in oversight. If true, then the difference in outcomes that we attribute to the variation in oversight may instead be the result of those other factors. Unfortunately, our analysis is mainly limited to the interrupted time-series approach for the Chicago analysis because of data limitations in other cities. However, to understand the impact of oversight and outrage on crime, we can compare Chicago to other major cities in the U.S. that were not affected by the change in Chicago oversight.

We are able to study how each event impacted crime rates in Chicago compared with other cities, as crime data exists in national datasets. Our main challenge is to construct three counterfactuals for Chicago around the time of each event.

For each event, we construct a balanced panel with $N = 65$ control cities and 46 time periods (months). Let y_{ct} be the outcome level for city c in period, where on the city of Chicago is treated. The treatment dummy P_{ct} equals one for Chicago the months after experiencing one of the events of interest and zero otherwise. In each panel, date $t_0 = 0$ corresponds to the date of the event with 30 pre-intervention (T_0) periods and 6 post-intervention periods (T^*) (including the treatment month).

We use synthetic control (SC) ([280, 228, 281]) and synthetic difference-in-differences (SDID) ([229]) methods to build the counterfactual Chicago. These approaches re-weight and match pre-event trends between Chicago and the control cities. The SC estimator that captures the average causal effect of an event, $\hat{\beta}^{sc}$, can be written as

$$(\hat{\beta}^{sc}, \hat{\mu}, \hat{\gamma}) = argmin_{\mu,\gamma,\beta} \sum_{c} \sum_{t} (y_{ct} - \mu - \gamma_t - P_{ct}\beta)^2 \hat{\omega}_c \tag{3.2}$$

where $\hat{\omega}_c$ are the weights for each cities that align pre-exposure trends in the outcome of control cities with those for Chicago, γ_t, is a time-fixed effect, and μ a constant. In contrast the SDID estimator allows for city fixed effects, α_c, and the time weights, λ_t, that balance pre-exposure time periods with post-exposure ones. Hence, the SDID estimator that captures the average causal effect of the event, $\hat{\beta}^{sdid}$, can be written as

$$(\hat{\beta}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\gamma}) = argmin_{\mu,\gamma,\alpha,\beta} \sum_c \sum_t (y_{ct} - \alpha_c - \mu - \gamma_t - P_{ct}\beta)^2 \hat{\omega}_c \hat{\lambda}_t \tag{3.3}$$

While the SC method re-weights treated and control units to match their pre-exposure trends, the SDID estimator assumes that unit and time weights exist such that the Chicago trends and the weighted average of the control units satisfy a parallel trends assumption. Finally, we follow [229] to compute the standard errors using a placebo analysis. We apply the SDID and SC model to 65 major cities that were not exposed to the events. [229] provides an algorithm that exploits placebo prediction using only the unexposed cities to estimate the noise level and then recovers the standard error of the estimates.

Compared with the ITS design, the SDID and SC designs recover estimates of the effect of each event in the long-run (15 months). Thus the point estimate will be the per-month cumulative effect of the event over the 15 months following the event's occurrence. Furthermore, whereas the ITS estimator is a before-and-after comparison, the SDID and SC estimates compare Chicago to a weighted combination of other large US cities as a counterfactual.

3.6 Results

3.6.1 The Effect of Outrage

We begin with the effect of outrage (Event 3) on outcomes. These results can be easily compared with other research using public scandals and viral videos which identify the effect of outrage on crime and civilian behavior. As previously discussed, Event 3, the public release of the video of Laquan McDonald's murder by CPD officer Jason Van Dyke, triggered a massive amount of news coverage, political changes, protests, and public criticism.

Crime Figure 3.2 displays the results for the effect of Event 3 on violent, property, and murder crime rates using both SC and SDID. The SDID results indicate that Event 3 caused an increase in violent crime by about 18% (relative to the mean) and specifically murder by 42% ($p < 0.01$), with a positive but imprecise effect on property crime. Table 3.2 displays the results for Event 3 in Column (3). Event 3 increased short-run violent and property crime significantly by 19% ($p = 0.01$) and 12% ($p = 0.01$) relative to the mean. Collectively, we can conclude that Event 3 caused an increase in crime in Chicago, consistent with prior literature on public scandals and crime and our framework relating to outrage.

Officer Behavior Next, Event 3 also had a significant effect on officer behavior. Table 3.2 shows that in the short-run arrests for violent and property crime increased by 7% ($p = 0.37$) and 8% ($p = 0.15$), which is not enough to match the increase in crime discussed above. Furthermore, discretionary or proactive officer activity declined significantly: arrests for drug crimes fell by 17% ($p = 0.11$) and other non-index crimes by 7% ($p = 0.33$). This decline in discretionary arrests is matched in a large decline in stops 73% ($p = 0.02$) and a decline in reported uses of force 14% ($p = 0.03$).[15] Again, this decline in officer activity is also consistent with both our predicted effect of outrage and the prior literature. However, following [282], we caution the reader that arrests may be a misleading measure of police productivity as they are mechanically related to the level of crimes. This is in line with [283] who argue that "arrests also signify a failure of prevention; if crimes are prevented in the first place, so are arrests and all of the ensuing costs of punishment."

Civilian Behavior Civilian behavior was also influenced by the scandal. At a public level, Event 3 increased articles about the CPD by 330% relative to the mean ($p = 0.01$), leaving little doubt about the magnitude of the public outrage and knowledge. However, Event 3 appears to have no effect on civilian reports of misconduct: the ITS result shows null effect of a 0.001% ($p = 1$) decline.

[15] Note that Event 3 also coincides with a policy change for how CPD officers had to fill out forms for stopping individuals. Shortly after Event 3, a reform on "stop and frisks" was enacted on January 1, 2016, after a settlement agreement between CPD and the ACLU. This reform requires officers to issue receipts if they search a person during a street stop.

However, in this case, the ITS assumption is violated as complaints are on a downward trend which is reverse upon Event 3 (see Figure 3.12), which would bias the result downward. Notably, this null effect is in the context of a large decrease in discretionary police-civilian contact and a decline in officer use of force. This is consistent with the our framework relating to the conflicting effects of outrage on complaints: while officers decrease contact with civilians and deliberate-misconduct in an attempt to avoid further scandal, allegations, and reduced morale, civilians' willingness to file complaints increases as they believe they are more likely to be believed.

Overall, these results indicate that following a major scandal, public outrage increased crime, decreased officer effort and proactive policing, and had a mixed effect on reported misconduct. These results are consistent with those of the scandal-based literature, however the reported misconduct result is consistent with civilian behavior changing significantly during a scandal. As we show, both police officers and civilians responded to outrage, so the results are unlikely to be driven by the effect of oversight alone.

3.6.2 The Effect of Oversight

Civilian Behavior In order to isolate the effect of oversight on outcomes separate from the effect of outrage, we use Events 1 and 2, which both were largely unknown to the public and also influence officers' perceived cost of receiving a complaint, as discussed in Section 3.4. The ITS results on *Tribune* articles is consistent with the second assumption, as shown in Table 3.2, Events 1 and 2 actually corresponded to decreases in articles about the CPD (though both have $p > 0.4$). Thus we can conclude that civilians were largely unaware of either of these events.

Based on the ITS results (columns (1) and (2) of Table 3.2), Events 1 and 2 both had large effects on complaints against police officers. Figure 3.12 shows that following Events 1 and 2 complaints against officers fell. As in Table 3.2, serious and FPS complaints fell by -31% ($p = 0.07$) and -0.51% ($p = 0.04$) after Event 1, while following Event 2 they fell by -40% ($p = 0.01$) and -27% ($p = 0.12$), respectively. In both cases, the events began a downward trend in all complaints, which persisted beyond the 6-month pre/post period used in the ITS, particularly for serious complaints. The ITS point estimates for Events 1 and 2, on all complaints are -84 and -69. For context, there were about 13,874 arrests, 41,427 stops, 30,866 reported crimes, and 485 uses of force per month in 2010.[16]

Given that the public was largely unaware of Events 1 and 2, a change in their propensity to report misconduct is unlikely. Rather, the evidence is consistent with officers reducing misconduct in order to avoid allegations due to anxiety over the future release of such complaints which were increased following each event.

Officer Behavior Naturally, if these drops in complaints corresponded with a reduction in error-misconduct only, we would expect to see a decline in officer effort or performance. However, as shown in Table 3.2, neither event had a statistically or economically significant effect on arrests for violent crime, and Event 1 is associated with an increase in arrests for property crime (though it is not statistically significant). For discretionary arrests relating to proactive policing, both events had no significant effect. These results are matched in other metrics for police activity, stops and use of force (stop data is not available for the time period around Event 1), with non-statistically significant and economically small effects of both Events 1 and 2. Despite the decline in complaints, the lack of change in officer activity metrics indicates that officers did not significantly reduce effort following these events.

Placing these results in our framework. We can conclude that the effect of Events 1 and 2 had a larger effect on deliberate-misconduct than on error-misconduct, as misconduct declined by officer activity was largely unaffected. Overall, these results contradict 'de-policing' theory and highlight the importance of considering misconduct beyond errors that result from crime-reduction effort.

Crime The most important effect, however, is the effect of oversight on crime. Based on the SDID results displayed in Figures 3.3 and 3.4, violent, property, and murder were not affected by either event in the long-run. These null effects are confirmed in the (short-run) ITS analysis in

[16]These events do not include all possible interactions which can produce a complaints, and stops and use of force are likely reported and many complaint originate in off-duty activity.

Table 3.2.

These results indicate that despite officers experiencing an increase in oversight, in that their expected cost of misconduct increased following each event, there was no effect on officer performance and overall no effect on crime rates. However, there was a substantial decline in complaints against officers. Given that civilians were largely unaware of the events, we believe this indicates that civilian behavior did not respond directly to these events and the complaint results are indicative of an effect on officer misconduct itself. Because the decline in misconduct was not accompanied by a net increase in crime nor a large decline in officer performance, the effect of oversight on error-misconduct is smaller than the effect on deliberate-misconduct. Thus, we find that oversight, unlike outrage, does not necessitate negative effects on officer effort or increases in crime.

Overall, these results are consistent with our framework and highlight the importance of distinguishing between error- and deliberate-misconduct and between oversight and outrage. However, they contradict the predictions of 'de-policing theory' and the usage of scandals as instruments for oversight in the existing literature.

3.7 Robustness

3.7.1 Bounding Treatment Effects

A fundamental challenge in studying the impact of outrage or oversight is that the outcomes of counterfactual policies are not observable. Our ITS analysis approach, or difference estimator, relies on a strong assumption that the outcome would have been constant in the months just before and after each event without the events occurring. However, in the presence of a counterfactual trend, our approach would not be valid. In the presence of a control group, a difference-in-differences (DID) approach would have enabled us to identify the impact of oversight and outrage under the assumption that Chicago and its counterfactual would have experienced in the absence of the similar change in complaints after the events.

In our context, we face several challenges similar to [226] and [227]. We provide a full description of the method in Appendix C.3. First, our events only treat Chicago, and the treatment is simultaneous across the whole city. Moreover, our sample is the population since our administrative data contain the universe of misconduct allegations filed against CPD officers. Thus, it is challenging to construct standard errors and confidence intervals since these measures of statistical precision rely on specifying a sampling data generating process that we do not feel confident to impose given our understanding of institutional details of the complaints process.[17]

As a result, we follow [227] to construct bounds of a DID estimate. We do not provide confidence intervals or standard errors for the reasons we discussed above. Finally, we discuss the sensitivity of the ITS estimates relative to the DID bounds. Note that the bounds identify the ambiguity created by the selection problem ([226]) and are not related to statistical imprecision created by sampling variability (i.e. confidence intervals or standard errors).

To construct the DID bounds, we consider the six months before the month of the event of interest as pre-period and five months after as post-period, as in the ITS analysis. Moreover, we use the data of the two years before each event as a control. For example, for the first two oversight events which occurred in July of their respective years (2009 and 2014), we use months around July 2007 and 2008 for the first event and July 2012 and 2013 for the second as controls. Similarly, the months around November 2013 and 2014 serve as control groups for the last event in November 2015. Finally, the treated groups and periods correspond to the data used for the ITS.

Figure 3.19 presents the impact of oversight and outrage on all complaints using the DID estimates and DID bounds from [226, 227]. We include the estimates from the ITS to compare the two approaches. Overall, the results from the bounding analysis are generally consistent with our findings from the ITS design. For Events 1 and 2, the bounds are consistently negative and relatively close with the ITS estimate lying between the upper and lower bounds. This suggests that the increase in oversight following Events 1 and 2 led to a reduction in the number of complaints. In contrast, there is substantial ambiguity about the impact of outrage on the number of complaints filed against police officers for Event 3. The upper bound is sometimes positive (the ITS estimate

[17] For instance, we discuss in Section 3.3 complaint data were altered by either CPD or the city in data releases after the complaints were initially made publicly available in late 2016.

is virtually zero). Furthermore, while the bounds were relatively tight for Events 1 and 2 (a few percentage point difference relative to the mean), for Event 3 the upper and lower bounds are generally separated by 100 percentage points. In other words, we cannot conclude that outrage increased or decreased police misconduct as the sign of the upper bound changes over time and the range of bounded effects is massive. Moreover, the bounds of the treatment effects are feasible for complaints, i.e., the lower bound is smaller than the upper bounds for all $t > 0$. As discussed previously, the bounds for the oversight events are consistently negative, suggesting a reduction in the number of complaints after increasing the level of oversight. Finally, DID estimates vary over time, which implies that treatment effects are heterogeneous. This result contrasts with the ITS estimates, which assume homogeneous treatment effects.

Similarly, we construct bounds for the crime results, as shown in Figure 3.20. Overall, the DID estimates and bounds are in line with the ITS results. Events 1 and 2 had generally negative or null effects on violent and property crime. Event 3 displays a parabolic shape for violent and property crime with bounds and estimates increasing in the first months then decreasing, though maintaining positive effect on both crime types. Similar to the results of the complaints, we find that the DID estimates are heterogeneous over time. Oversight led to a reduction in both property and violent crimes, and the ITS estimates tend to be within the bounds. However, the outrage event led to an increase in property crimes. Under the model assumption, the bounds for violent crimes are indeterminate for the outrage event (i.e., the bounds are infeasible).

3.7.2 Reporting Suppression

A natural concern would be that in response to Events 1 and 2, officers or the CPD may increase suppression of complaint allegations. This could be in the form of officer-level suppression (threatening would be complainants) or at a departmental level (pressuring complaint-takers to ignore or mis-file complaints). While this is certainly possible, we believe it is unlikely to be driving the entire decline following each of these events. As discussed in the DOJ report on the CPD in 2017, much of the suppression of complaints came from institutional barriers that prevented the investigation of, not the reporting of, complaints [284]. First, as shown in Table 3.3, the share of complaints filed by officers actually increased (17% (p = 0.36) and 50% (p = 0.03) relative to the mean share), meaning the decline is entirely driven by civilian-filed complaints. If there was an organized effort to suppress complaints, we would likely see a change in internal complaints as well as civilian ones. Second, complaints can be filed over the phone or email, but they are only investigated if the complainant signs an affidavit. If the share of no affidavit complaints rose significantly, we may suspect a large amount of suppression– however there is not a significant increase in the share of no affidavit complaints, as shown in Table 3.3.

3.7.3 Robustness of Synthetic Difference-in-Differences Estimates

In order to test the robustness of our national-level results, we use SC and different samples of cities. Figure 3.18 displays the results. Overall, the SDID results are consistent across samples using the main sample, large cities only ($\geq 500,000$ population), and large cities with no investigations by the DOJ to verify that our results are not sensitive to more restrictive sets of control cities. Using SC, however, reduces the magnitude of our results on violent crime across samples. However, both the SC and SDID results are consistent and highly similar across samples in terms of the effect on murder. Furthermore, in no robustness check does either Event 1 or 2 have a statistically significant effect on murder, violent, or property crime. Overall, we can conclude that Event 3 (outrage) caused an increase in murder in Chicago, while both oversight events did not cause significant increase in crime.

Relatedly, because we use 30 pre-treatment and 15 post-treatment periods, the pre-period for Event 3 will overlap with Event 2. To ensure this is not contaminating our results, we redo our SDID and SC analysis with symmetric time windows (15 months pre and post including the event month) such that Event 3's pre-period does not overlap with Event 2. The results for both SC and SDID are shown in Figure 3.18, and they are very similar to the main results.

3.7.4 Ferguson Timing

A concern with Event 2 in particular is that it occurred around the same time as the Ferguson protests following the killing of Michael Brown on August 9th, 2014. However, we believe this is unlikely to be an issue, and if it did have an effect on officers and/or civilians, it would actually cause an increase in crime and complaints, which we do not find in our results. First, suppose that Ferguson only influenced CPD officers, and actually acted as an oversight event. This is unlikely because while the killing occurred in August 9th, the first mentions of Ferguson on SCC appeared in late August, and thus the large decline in complaints in August is unlikely to be driven by an event which only partially treated the month. Second, officers were much more likely to be focused on the local complaint and FOIA issues around Event 2 which fully treated the month of August (the union issued its response in late July), and the union was clearly focused more on Event 2. The August 2014 newsletter does not mention Ferguson at all, while the September 2014 newsletter mentions Ferguson on page 9 (of 12) but discusses FOIA and complaints on the front page. Overall, it is unlikely that Ferguson, which is now considered the beginning of the police oversight movement in the US, actually registered significantly with CPD members at the time who were already under significant local pressure and pre-existing lawsuits and court orders.

Second, Ferguson would have likely caused more of a public response than an officer-only one. Note that there was only a small increase in discussions of police misconduct in August of 2014 which quickly declines back to baseline in Chicago (see Panel C of Figure 3.1). If the public did respond, we would expect to see the decline in complaints be attenuated due to increased civilian reporting. Furthermore, we would expect that Ferguson would have caused a smaller scale outrage event, and thus we would expect to see an increase in crime (due to officers pulling back, civilians engaging less, increased legal cynicism and social disorder), which we do not see in the results. Overall, it is unlikely that Ferguson confounds our main results for Chicago: that Event 2 caused a decrease in complaints and no increase in crime through increased oversight.

3.7.5 Officer Level Estimates

Though the treatments are at the city-level, we are interested in how individual officer outcomes changed, providing us with an additional set of point estimates. We estimate the change in officer (i) outcome Y following each event in a month (t) using an officer-month level panel, after controlling for officer, unit, year, and month fixed effects:

$$Y_{it} = \beta_1 1\{T_1 < t\} + \beta_2 1\{T_2 < t\} + \beta_3 1\{T_3 < t\} + \gamma_i + \lambda_{Y(t),M(t)} + \psi_{Unit(i,t)} + \epsilon_{it} \tag{3.4}$$

where T_j indicates the month of event j, and β_j is the effect on the officer being in the period following event j– the other variables are fixed effects for month, year, and unit.

We provide results using an officer-level approach. Notably, the complaints used as an outcome are any complaint with a known officer (no restrictions on having a complainant) and the same complaint counts for multiple officers. The results are shown in Table 3.4, Events 1 and 2 caused a 11% ($p < 0.01$) and 5% ($p = 0.02$) decline in complaints based on the officer-level analysis. Similar to the ITS results, Event 1's effect is larger for FPS complaints, while Event 2's is larger for serious complaints. In other specifications, the smallest and most noisy point estimates are when including complaints with missing complainants (as in the officer-level regressions)– Events 1 and 2 display large declines 18% ($p = 0.01$) and 11% ($p = 0.26$) (see Table 3.3).

The officer-level results also provide more information on changes in officer activity. The results for Events 1 and 2 in Table 3.4 show a small decline in violent and property arrests, but Event 2 has a large effect on drug arrests (a 17% decline) and Event 1 displays an increase of 6%. Taking into account use of force, Events 1 and 2 have a consistent negative effect on complaints and not other officer activity, particularly across other specifications. Event 3, at the officer level, displays an increase in allegations of misconduct, for both serious (8%) and FPS complaints (29%) (both $p < 0.01$). Furthermore, there is a decline in arrests (specifically violent and drug) and use of force. These results are generally consistent with those of the main ITS analysis.

3.7.6 Changing T

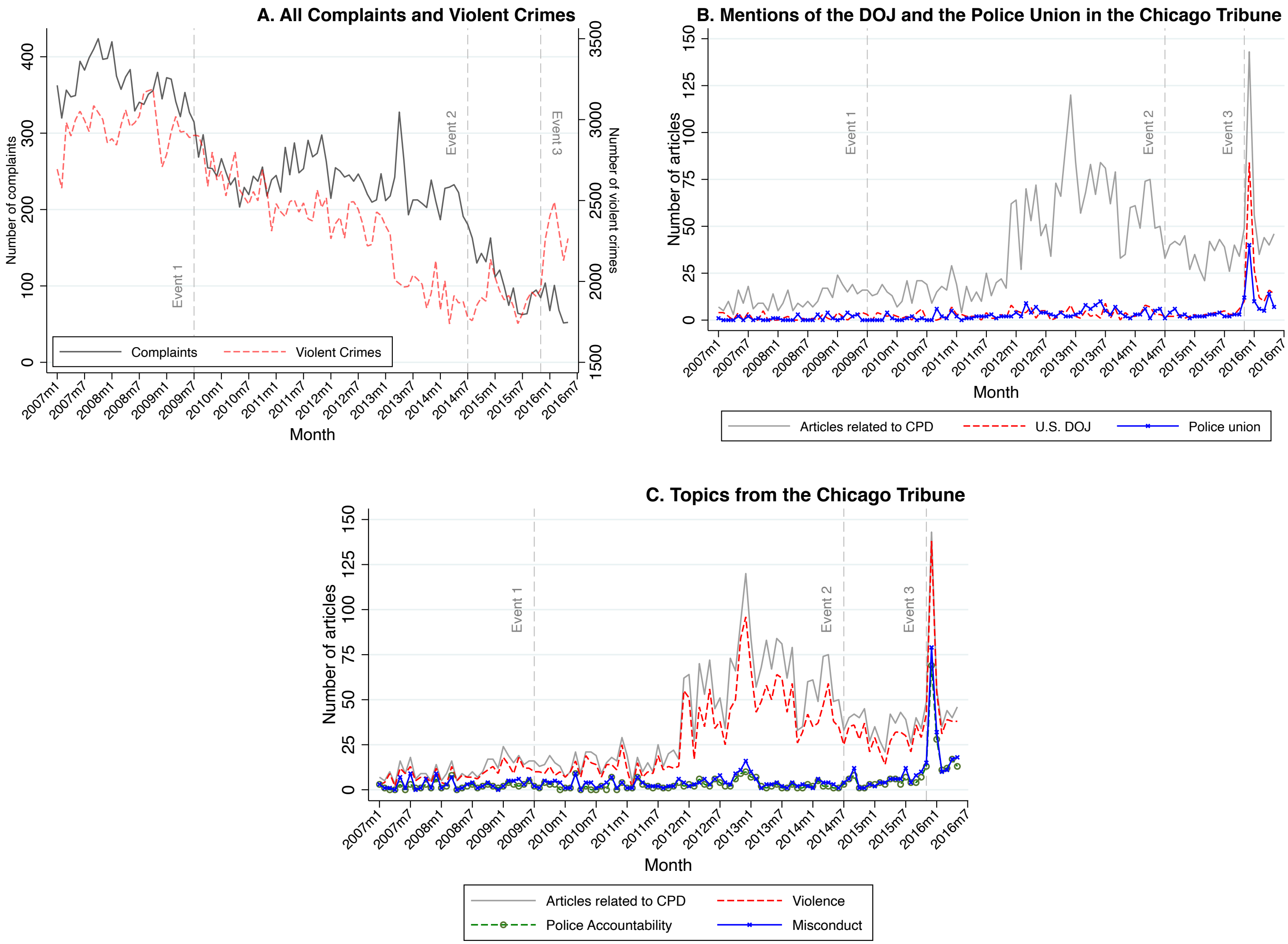
We also want to ensure our results are robust to changing the time span for the pre- and post-periods. Table 3.5 displays the results for changing the time-span in our ITS analysis to $T = 4$ and $T = 8$. Overall, these results are consistent with the main results using $T = 6$.

3.8 Conclusion

In this paper, we provide novel evidence for the effect of oversight on misconduct, policing, and crime. This differs greatly from previous work on police oversight because we identify events which solely increased oversight (the effective cost of misconduct) and did not coincide with public outrage. In this vein, we provide a conceptual framework which illustrates the difference between oversight and outrage. As we discuss, while outrage has a predictable effect on crime, it is not clear ex-ante that oversight will increase crime as misconduct may be due to deliberate and counterproductive actions as well as errors resulting from officer effort, and so reducing deliberate-misconduct will not increase crime while reducing error-misconduct can increase crime. Furthermore, we note that outrage should have a mixed effect on reported misconduct (an outcome not previously studied in the scandal-based literature), as officers may reduce misconduct during a period of outrage but civilians will be more likely to report misconduct.

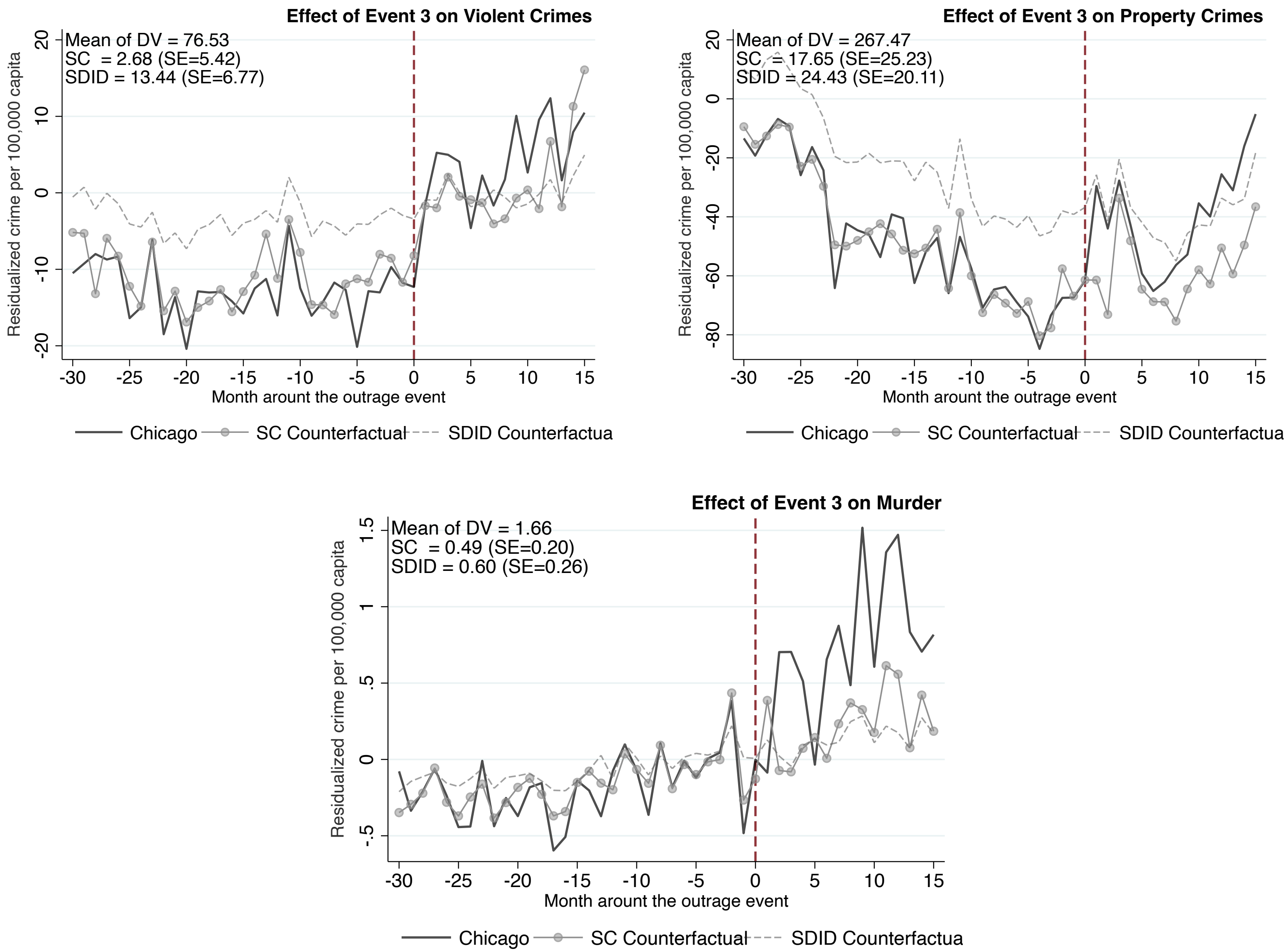
We provide empirical evidence for this framework which clearly distinguishes between oversight and outrage. We find that following a public scandal, crime, particularly murder, increases while officers reduce effort (e.g., discretionary arrests and stops) and negative interactions with civilians (uses of force). Nevertheless, reported misconduct is either unaffected or increases: civilians are more likely to report perceived misconduct while police-civilian contacts decline. In contrast, we show that increases in police oversight, specifically two events relating to the public release of officer misconduct records, did not increase crime. But, as the social and reputational costs of a public release of misconduct accusations meant the cost of misconduct increased, officers reduced their misconduct, as measured by complaints. We attribute this to a decline in misconduct and not a decline in reporting because while these events were highly salient for officers, the public was largely unaware.

Overall, this paper demonstrates that outrage is not oversight, and conflating the two can lead to erroneous conclusions about the costs and benefits of police oversight. Furthermore, we find evidence that oversight alone can reduce misconduct and not lead to an increase in crime or significant changes in policing outcomes.



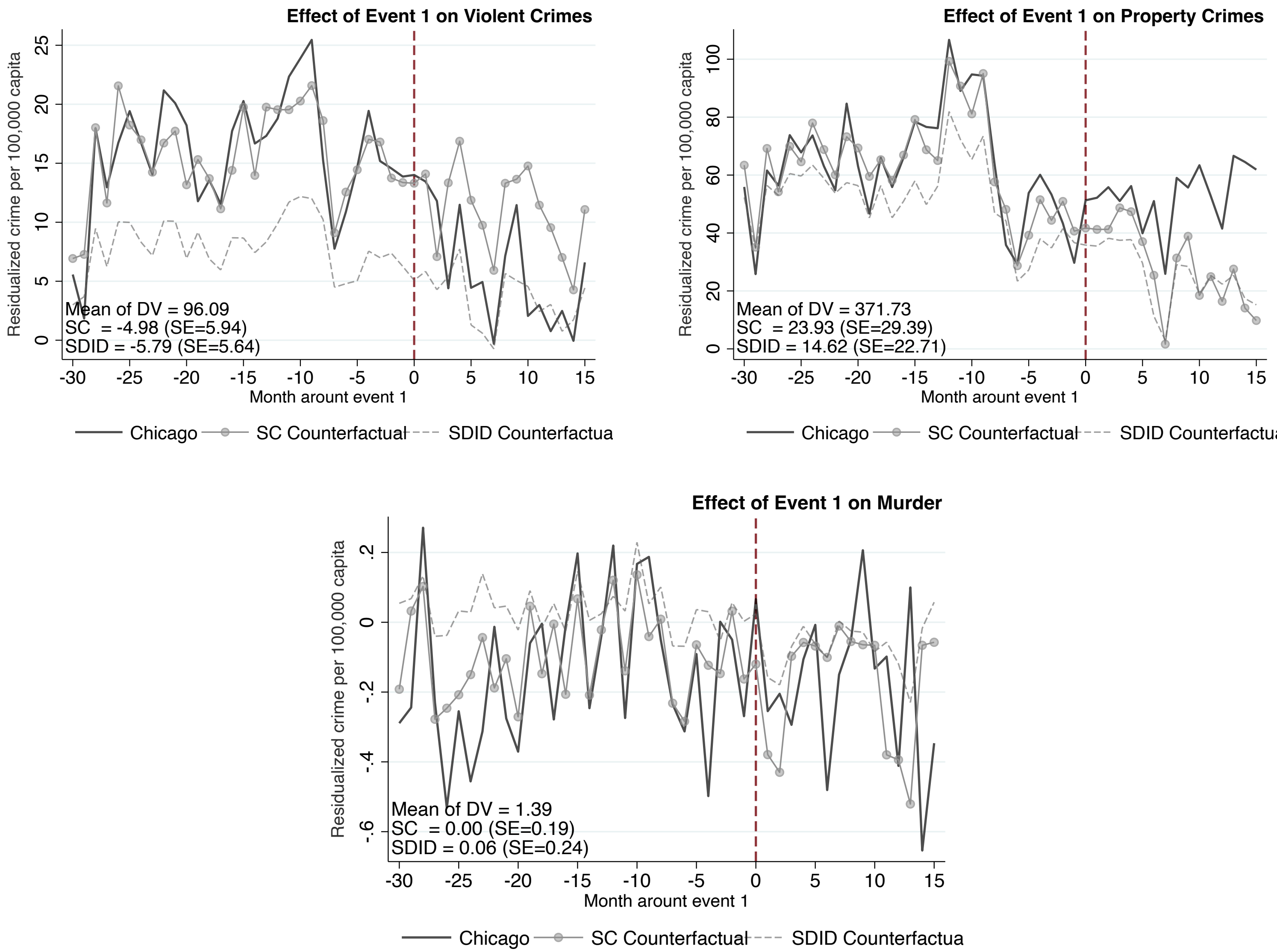
Notes: Figure A presents the total number of misconduct allegations filed against a Chicago police officer and the number of violent crimes from January 2007 to May 2016. Figures B and C present the number of articles mentioning the Chicago Police Department in the Chicago Tribune from January 2007 to May 2016. Figure B presents the frequency distribution of articles mentioning the DOJ and the Chicago police union. Figure C shows the frequency distribution of articles related to gun violence, police accountability, and police misconduct. The vertical dashed lines represent three events that have plausibly varied the level of oversight against the Chicago Police Department between 2007 and 2016. We account for seasonal patterns in complaints and violent crimes by removing the month effect. Section 3.4, in the main text, describes each event in detail. Violent crimes correspond to murder, assault, and robbery.

Figure 3.1: Time-series of Complaints, Crime, and News Coverage



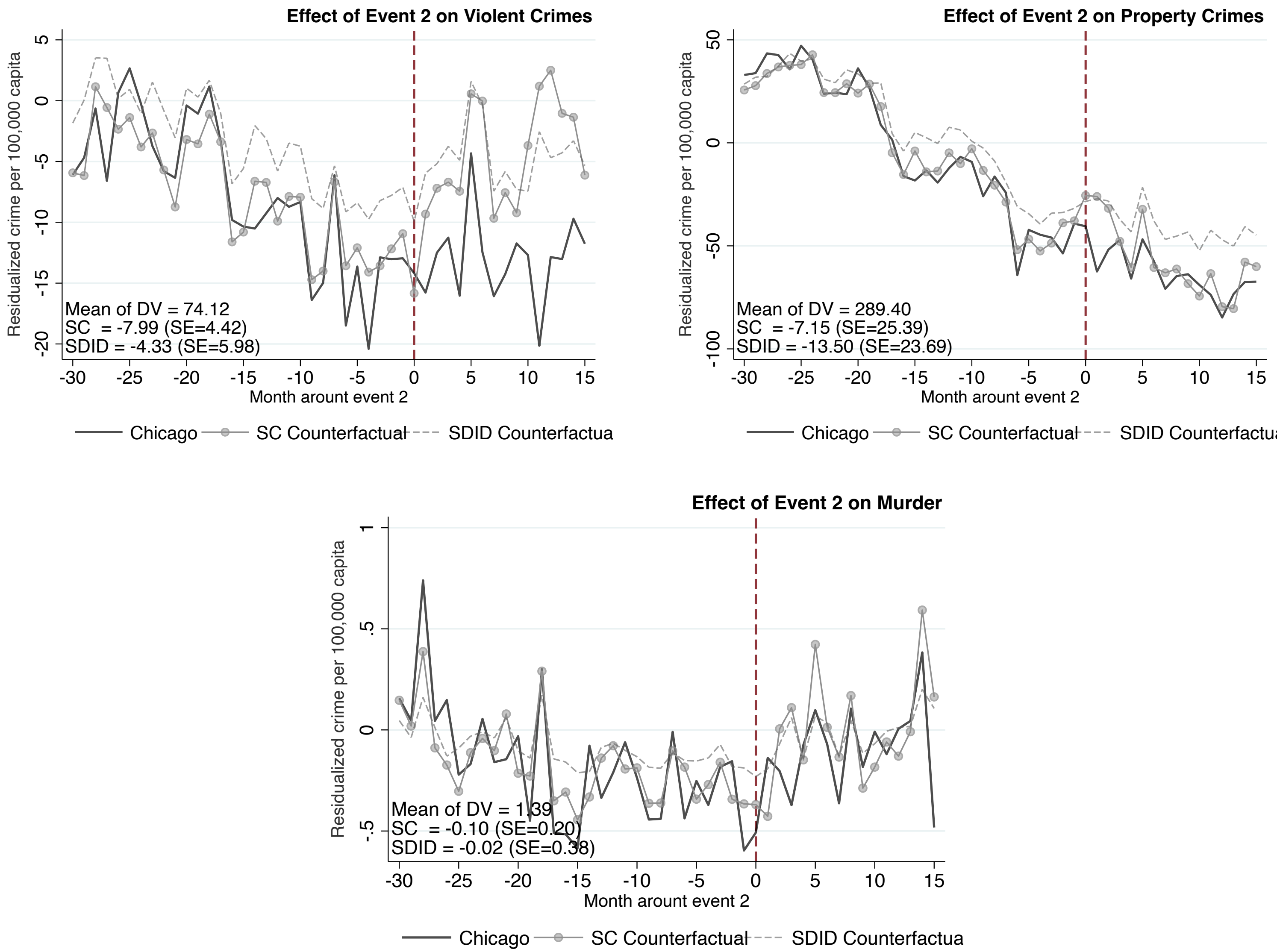
Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.2: Effect of Outrage on Crimes



Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.3: Effect of Oversight on Crimes (Event 1)



Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.4: Effect of Oversight on Crimes (Event 2)

Complaint Type	Complaint Description	Total N.	Share	Cumulative Share	Po Filed	Sustained	Has Complainant
FPS	10U-INADEQUATE/FAILURE TO PROVIDE SERVICE	12658	0.162	0.162	0.025	0.015	0.904
UFV	05A-ARRESTEE - DURING ARREST	7929	0.101	0.263	0.021	0.015	0.937
Search	03G-MISCELLANEOUS	6704	0.086	0.349	0.010	0.005	0.916
Search	03D-ILLEGAL ARREST	6500	0.083	0.432	0.009	0.003	0.908
Search	03C-SEARCH OF PREMISE/VEHICLE WITHOUT WARRANT	6248	0.080	0.512	0.016	0.003	0.938
Other	10Z-MISCELLANEOUS	4057	0.052	0.563	0.165	0.073	0.940
FPS	10J-NEGLECT OF DUTY/CONDUCT UNBECOMING - ON DUTY	3336	0.043	0.606	0.416	0.231	0.915
ALU	04E-PRISONER'S PROPERTY - INVENTORY/RECEIPT	2918	0.037	0.643	0.035	0.021	0.887
Search	03B-SEARCH OF PERSON WITHOUT WARRANT	2910	0.037	0.681	0.028	0.000	0.910
UFV	05B-ARRESTEE - AFTER ARREST, PRIOR TO LOCKUP	2412	0.031	0.711	0.021	0.012	0.926
Other	Missing	1750	0.022	0.734	0.218	0.000	0.714
Other	10V-INVENTORY PROCEDURES	1608	0.021	0.754	0.203	0.093	0.974
UFV	05L-UNNECESSARY PHYSICAL CONTACT - ON DUTY	1251	0.016	0.770	0.058	0.026	0.920
Other	10T-REPORTS - FAILED TO SUBMIT/IMPROPER	1197	0.015	0.786	0.206	0.096	0.911
UFV	05Q-CIVIL SUIT - THIRD PARTY	1173	0.015	0.801	0.000	0.003	0.990
Other	10D-COMMUNICATION OPERATIONS PROCEDURES	1035	0.013	0.814	0.007	0.007	0.848
UFV	01B-RACIAL/ETHNIC, ETC.	920	0.012	0.826	0.028	0.027	0.939
UFV	05K-DOMESTIC ALTERCATION/INCIDENT - OFF DUTY	871	0.011	0.837	0.188	0.160	0.916
Other	09J-MISCELLANEOUS	867	0.011	0.848	0.489	0.273	0.926
UFV	05M-UNNECESSARY PHYSICAL CONTACT - OFF DUTY	767	0.010	0.858	0.037	0.023	0.930

Table 3.1: Composition of 20 Most Common Complaints

Note: Table contains summary statistics for the 20 most common types of complaints with incident dates between 2007 and May 2016. Po Filed indicates the share of those complaints which a police officer filed the complaint (otherwise a civilian filed). Sustained refers to the final finding of the complaint being either sustained or marked as 'disciplined'. And having a complainant means at least one complainant was recorded for the complaint.

Outcome	Effect of Event (% Change Relative to Mean)		
	Oversight (1)	Oversight (2)	Outrage (3)
Complaints			
All	-0.342 (p = 0.01)	-0.281 (p = 0.03)	-0.001 (p = 1)
Serious	-0.312 (p = 0.07)	-0.401 (p = 0.01)	0.005 (p = 0.99)
FPS	-0.517 (p = 0.04)	-0.274 (p = 0.12)	0.037 (p = 0.78)
Other	-0.251 (p = 0.23)	-0.029 (p = 0.93)	-0.048 (p = 0.82)
Public			
Tribune Articles Related to CPD	-0.444 (p = 0.45)	-0.009 (p = 0.99)	3.308 (p = 0.01)
Crime			
Violent	-0.067 (p = 0.18)	0.014 (p = 0.82)	0.191 (p = 0.01)
Property	0.01 (p = 0.81)	-0.057 (p = 0.3)	0.123 (p = 0.01)
Arrests/Stops/Force			
Violent	0.008 (p = 0.9)	0.043 (p = 0.59)	0.065 (p = 0.37)
Property	0.081 (p = 0.18)	-0.019 (p = 0.88)	0.083 (p = 0.15)
Drug	0.007 (p = 0.92)	-0.036 (p = 0.81)	-0.172 (p = 0.11)
Other	-0.02 (p = 0.78)	-0.001 (p = 1)	-0.073 (p = 0.33)
Stops		-0.147 (p = 0.35)	-0.733 (p = 0.02)
Force	0.006 (p = 0.95)	0.006 (p = 0.93)	-0.137 (p = 0.03)

Table 3.2: Effect of Oversight and Outrage on Outcomes

Note: Table contains the ITS results from estimating equation 3.1 with 6 pre-event months and 5 post-event months for each outcome. Each column refers to an event. Stop data could not be obtained pre-2010, so the effect of Event 1 on stops could not be estimated. Dependent variables are residualized by month to remove seasonality. Coefficients are all scaled as a share relative to the mean of the dependent variable in the 6-months before the event. P-values are computed by ranking the effects the 3 events along with the effects of 70 placebo periods, which exclude the period of the events and the events themselves.

Outcome	Effect of Event (% Change Relative to Mean) Oversight (1)	Oversight (2)	Outrage (3)
Share Officer Filed	0.167 (p = 0.36)	0.5 (p = 0.03)	0.889 (p = 0.01)
Share No Affidavit	-0.034 (p = 0.53)	0.017 (p = 0.95)	0.22 (p = 0.01)
All Complaints (Including Missing Complainants)	-0.182 (p = 0.01)	-0.114 (p = 0.26)	-0.041 (p = 0.75)
White Complainants	-0.298 (p = 0.07)	-0.088 (p = 0.67)	0.033 (p = 0.88)
Black Complainants	-0.372 (p = 0.01)	-0.345 (p = 0.03)	-0.02 (p = 0.9)

Table 3.3: Additional Complaint Results

Note: Table contains additional ITS results from estimating equation 3.1 with 6 pre-event months and 5 post-event months for each outcome. Each column refers to an event. Share officer filed refers to the share of complaints filed by an officer, not civilian; share no affidavit refers to the share of complaints without a signed affidavit; all complaints includes complaints with missing complainants; white and Black complaints are categorized based on reported race in the complainant data. Dependent variables are residualized by month to remove seasonality. Coefficients are all scaled as a share relative to the mean of the dependent variable in the 6-months before the event. P-values are computed by ranking the effects the 3 events along with the effects of 70 placebo periods, which exclude the period of the events and the events themselves.

	Effect of Event (% Change Relative to Mean)		
Outcome	Oversight (1)	Oversight (2)	Outrage (3)
All Complaints	-0.105 (se = 0.027) (p = 0)	-0.053 (se = 0.023) (p = 0.02)	0.139 (se = 0.025) (p = 0)
Serious Complaint	-0.125 (se = 0.037) (p = 0.001)	-0.096 (se = 0.03) (p = 0.001)	0.082 (se = 0.029) (p = 0.004)
FPS Complaint	-0.183 (se = 0.052) (p = 0)	-0.061 (se = 0.047) (p = 0.192)	0.289 (se = 0.061) (p = 0)
Violent Arrests	-0.014 (se = 0.007) (p = 0.036)	-0.038 (se = 0.023) (p = 0.099)	-0.174 (se = 0.026) (p = 0)
Property Arrests	-0.053 (se = 0.009) (p = 0)	-0.035 (se = 0.031) (p = 0.26)	-0.031 (se = 0.033) (p = 0.351)
Drug Arrests	0.055 (se = 0.009) (p = 0)	-0.172 (se = 0.031) (p = 0)	-0.248 (se = 0.027) (p = 0)
Other Arrests	0.086 (se = 0.006) (p = 0)	-0.042 (se = 0.017) (p = 0.015)	-0.08 (se = 0.017) (p = 0)
Use of Force	0.039 (se = 0.029) (p = 0.177)	-0.045 (se = 0.029) (p = 0.128)	-0.108 (se = 0.034) (p = 0.001)

Table 3.4: Officer-Level Results

Note: Table contains results from estimating equation 3.4 using officer-month level observations between 2009 and May 2016. The point estimates can be seen as the additional effect of each event, given the prior events occurring. Fixed effects include (additively) officer, month, year, and unit fixed effects. Standard errors are clustered at the officer level.

10/23/21, 12:32 PM Second City Cop: Danger Will Robinson! DANGER!!!

More Create Blog Sign In

SECOND CITY COP

SARCASM AND SILLINESS FROM A WINDY CITY COP

FRIDAY, JULY 24, 2009

Danger Will Robinson! DANGER!!!

This has the potential for being bad. Very bad:

- *In a case some argue could throw open the long-standing secrecy behind police internal investigations, the 4th District Court of Appeals in Springfield has ruled internal affairs files are a public record regardless of the outcome of the probe.*

  Attorneys specializing in Illinois public records law said Thursday it is the first such ruling of its kind in the state and therefore binding on trial courts statewide. It could also have repercussions for long-running complaints about Chicago police brutality.

Great. Compelled testimony rears its ugly head again. Remember, you have less rights in an administrative investigation than a criminal one. You are ordered to answer questions. You are ordered to testify give information against your own self interests. You are assumed to be in possession of facts and recollections of incidents merely by your presence on a scene.

And if a single error or omission can be attributed to anything you put on paper, you will be fired.

Not only that, the bottom-feeders are going to use this civilly to ruin cops across the board. Anything you own, anything you've worked for, anything you've accomplished by dint of hard work serving and protecting is going to have to be sold, pawned or given away to defend yourself in a Crook County courtroom with a juries who already give felons millions.

- *The Sangamon County Sheriff and the State Attorney's Appellate Prosecutor—who argued the case—have 35 days to decide whether to appeal the ruling to the Illinois Supreme Court*

The Local and State FOP better get on the ball here. The National Lodge might not be a bad idea either.

CONTRIBUTORS

SCC

SCC (RETIRED)

secondcitycop.blogspot.com
sarcasm and silliness from a second city cop
THE SOURCE for Chicago Police info
Opinions. Observations. Ornery Human Beings.

POLICE LIMIT
THE FIRST CLUSTER
THE FIRST COLLECTION OF POLICE LIMIT CARTOONS BY GAREY MCKEE!
BUY NOW!

MOST VALUABLE
CHICAGO
BLOGGER

https://web.archive.org/web/20200613102108/http://secondcitycop.blogspot.com/2009/07/danger-will-robinson-danger.html 1/20

Notes: This figure presents a comment from Second City Cop posted in July 2009. The post is related to event 1, the court ruling in *Gekas v. Williamson* indicated that "internal affairs files are a public record regardless of the outcome of the probe." SCC is a (now offline) forum for CPD officers to anonymously discuss work, news, and drama within the department without a filter.

Figure 3.5: Second City Cop Post related to Event 1

November 2009 Page 3

Second Vice-President's Report, *By Frank DiMaria*

It's Just a Witness Report Officer

The Lodge must receive at least 50 calls a month from our Members telling us that they were notified by either IAD or IPRA because they are being required to submit either a formal statement or a written report and that they are not accused but only a witness. The first comment from these officers sometimes will be that I am just letting the Lodge know but I don't think I need a lawyer, because I am only a witness regarding a CR investigation. The best advice I can always give to these Officers is that you are entitled to representation pursuant to Section 6.2 of our Contract which describes and details how the Department or IPRA will conduct Witness Statements involving our members whether written or oral. The affected Officer upon being notified by a supervisor or on Department computer that they are a witness in a CR investigation have the right to request to secure legal counsel. The Lodge strongly recommends that any time you must respond to this type of notification you contact the Lodge and we will provide an attorney to assist and represent you during the interview. Officers in the event that a report is required in this investigation notify the Lodge and we will assist you in the completion of your written report. Officers don't take anything for granted in regards to these investigations. You could be a witness on Monday and then the accused on Tuesday. When notified of being a witness in a CR investigation, please contact the Lodge and we will assist you in these matters. It is your contractual right!

Parole Board Hearings

The Lodge was recently notified of the Parole Board Hearing which were conducted at 26th & California on 21 October 2009 for us to protest the possibility for parole for the convicted murderer of PO Herman Stallworth, #10965, 003rd District, End of Watch 24 May 1967.

Attending this hearing was a representative of the Department, the Police Memorial Foundation, an Assistant States Attorney (ASA), two CPD Detectives and I, representing the Lodge. The hearing which is video taped and was presided over by a member of the Illinois Prisoner Review Board who would receive our testimony and evidence. The ASA spoke to the member of the prisoner review board and retold the story of how PO Stallworth was murdered and his partner PO Ervin was shot on the night of 24 May 1967 while conducting a routine traffic stop for no other reason that they were doing their duty serving and protecting the citizens of the city of Chicago and how this coward only surrendered after he ran out of bullets in the handgun which he fired at responding police officers. The letters from the surviving widow and partner were also read and detailed how their lives had been impacted, changed forever by this violent act and that this murderer should not be released. We the representatives in attendance then all read our objections into the record asking that the Parole Board not release this convicted murderer before he complete his lawful sentence. The En Banc Hearing in this case as well as 2 other cases involving the convicted murderers of Chicago Police Officers will be heard at the State Capitol in Springfield, Illinois on 19 November 2009. We must remind those who are responsible for ensuring that these dangerous individuals never return to society.

We will travel by bus with other police officers and supporters to make our presence known and object to any possibility of parole for these convicted cop killers before their lawfully imposed sentenced is completed.

Notes: This figure presents an article from the FOP Lodge 7 newsletter dated November 2009. The article is from the Second Vice President of the union and warned officers of the severity and potential risks of being named in a complaint investigation.

Figure 3.6: Newsletter November 2009

President's Report, *by Dean C. Angelo, Sr.*

Release of CR Files (Kalven Court Decision)

A recent Appellate Court decision concerning the release of CR files pursuant to Freedom of Information Act (FOIA) requests as part of *Kalven v. the City of Chicago & the Chicago Police Department* has become a major concern to the Membership; and rightfully so. When this litigation began back in 2009, the Lodge did not formally join the City or the Department in this legal challenge, thus leaving the City and the Department to challenge the FOIA requests to release CRs and RLs (Repeater Lists) on their own. Now, after the decision has been heard and appealed, the Lodge has no legal standing or litigant position to file any additional actions.

So what does this mean to the Membership? In compliance with the court decision, lists of past complaints against Chicago Police Officers will now be released to the public for review. We initially spoke about this case at our first General Meeting in April and mentioned then that this decision would wind up being something that every working officer would have an extensive interest in. We contacted the Lodge's attorneys and requested their take on the impact. (Further information that details the attorney's input is available on page six of this Newsletter).

More History

In December 2009, (Attorney) Kalven filed suit against the City (and the CPD) under the Freedom of Information Act (FOIA), to obtain Complaint Register (CR) files as well as Repeater Lists (RL). Initially, the Circuit Court held that Kalven was entitled to the RLs, but not to the CRs. After an appeal, the Court ruled Kalven was entitled to both the RLs and CRs, (subject to complainant and witness redaction) .

The intention of the Freedom of Information Act (as argued in Kalven), was to allow the public to review the actions of public bodies to ensure full disclosure of public employees and make them available for review; such as with Chicago Police Officers. An individual working for a private employer would not face this level of scrutiny. As we are all well aware, a CR Investigation of a Police Officer does not establish any wrongdoing on the part of the Officer; it is a fact that most allegations are either Not Sustained, Unfounded or, Exonerated.

An additional concern of ours is the impact this type of disclosure might have on the reputation of the individual Police Officers and quite frankly, the Chicago Police Department as well. A Police Officer might suffer undue scrutiny from family members, friends, neighbors and even co-workers, when this information is released. Furthermore, an additional risk now exists that the public (including some criminal defense attorneys) can start to access this information through a Freedom of Information Act request in a much more exaggerated manner.

Lodge Position

The Lodge emphatically believes the Court erred in its ruling. Consequently, we believe the statute should be amended to exempt both CRs and RLs from public view and production. The Lodge sees this decision as unfair prejudice and that such disclosures may have a negative impact on individual Police Officers. There also exists a likelihood that such public exposure will discourage many intelligent, skilled and dedicated individuals from continuing, or even beginning a career in public service. When is enough, truly enough?

A Final Update on Kalven v. City

Although we cannot alter the Court's decision concerning the release of the CR and RL lists, we might have one option available to possibly challenge some of the names on those lists. On 28 July 14, the Lodge filed an Emergency Motion requesting a Temporary Restraining Order

(Continued on page 2)

Notes: This figure presents an article from the FOP Lodge 7 newsletter dated August 2014. The article is from the President of the union and discusses the outcome of the court ruling, *Kalven v. Chicago*: "In compliance with the court decision, lists of past complaints against Chicago Police Officers will now be released for public review." The article was front the page of the newsletter.

Figure 3.7: Newsletter August 2014

September

2014

F.O.P. News

Official Publication of Chicago Lodge No. 7 -
1412 W. Washington Blvd. Chicago, IL 60607-1821 Phone: 312-733-7776 FAX: 312-733-1367
www.chicagofop.org

Fraternal Order of Police
Chicago Lodge 7
1412 W. Washington Blvd.
Chicago, IL 60607

Inside this issue:

President's Report, *by Dean C. Angelo, Sr.*

Contract Update

As a way to continue to keep the Members as informed as possible, I would like to bring everyone up to speed on where we are at, Contract-wise. Right now we are meeting in closed sessions that are only made up of 3 to 4 Lodge Representatives and 3 to 4 City Representatives. These smaller grouped sessions have been taking place for the past few weeks and could be considered the "Core of the Core". Needless to say, we are still moving through several of the issues that remain, but are now doing so at a much quicker pace. We continue to be hopeful that we will be able to bring something for consideration to the full Board of Directors and then to the Unit Reps.

Although we have made strides in many areas, there remain several key issues that need to be discussed and agreed upon, but the overall environment within the room is one that offers open rhetoric and dialogue that is focused on resolve…most of the time. Just so that we are all very clear on this process, the negotiations are continuing and no final agreement has been reached. Furthermore, and completely up front, if several of our remaining (and critical) issues are not taken seriously, we always have other options available to us including Arbitration. It is not the Lodge's intention to seek out Arbitration, but even though we are in a good place communication-wise, we will not forget nor take likely our purpose and responsibility to each and every one of you.

CRs and FOIA

For the past several weeks now we have been consistently speaking to the Membership on the particulars concerning the Freedom of Information Act requests that seem to have taken on a life of their own. As a matter of information, other than one minimal at best consideration, there are no formal requirements or formats that need to be adhered when submitting a FOIA request to the City or the Chicago Police Department. Any form of "written" request will suffice in meeting the present minimal standard of compliance, FOIA-wise. The only requirement necessary to meet this low bar process is that the request be typed, handwritten or even in email form. We find this totally unacceptable and will work on how to best attack this oversight by the courts during the next legislative session in Springfield.

Our attorneys have already been working on new language to address what they believe will put us in the best light to alter what is presently in place. As a way to garner more support behind this fight, we now have the ability to depend on and utilize the Illinois State Lodge, Troopers Lodge 41 and the Illinois Labor Council to assist us in this battle. This multi-pronged approach is something Lodge 7 has not had in its arsenal for several years, but it is now readily available to us and we plan on taking advantage of it at every opportunity. We also intend to not only build this upon the strengths of this relationship, but to have it become an ongoing collective force; a force that will be difficult to ignore and one that will require accommodations. As with most everything of importance, there is strength in numbers.

Outside Deployments…Really?

As everyone is well aware, there have been several recent deployments of outside law enforcement personnel from multiple agencies assigned to work in some of the higher crime areas throughout the City. First off, Cook County Sherriff Officers have been regularly working in a couple of the west side districts for the past several weeks. They have almost become what some of our Members are saying, a "regular thing" now.

Follow those assignments with the more recent influx of Illinois State Troopers who are now being deployed to several of the CPD Fugitive Apprehension Teams. The plan is to place two Troopers with the present five member teams, thus increasing the manpower by an additional couple of bodies each. The Troopers are being housed just west of the City limits so that they don't have to travel the distance from their rural homes to the new assignments in the Chicago.

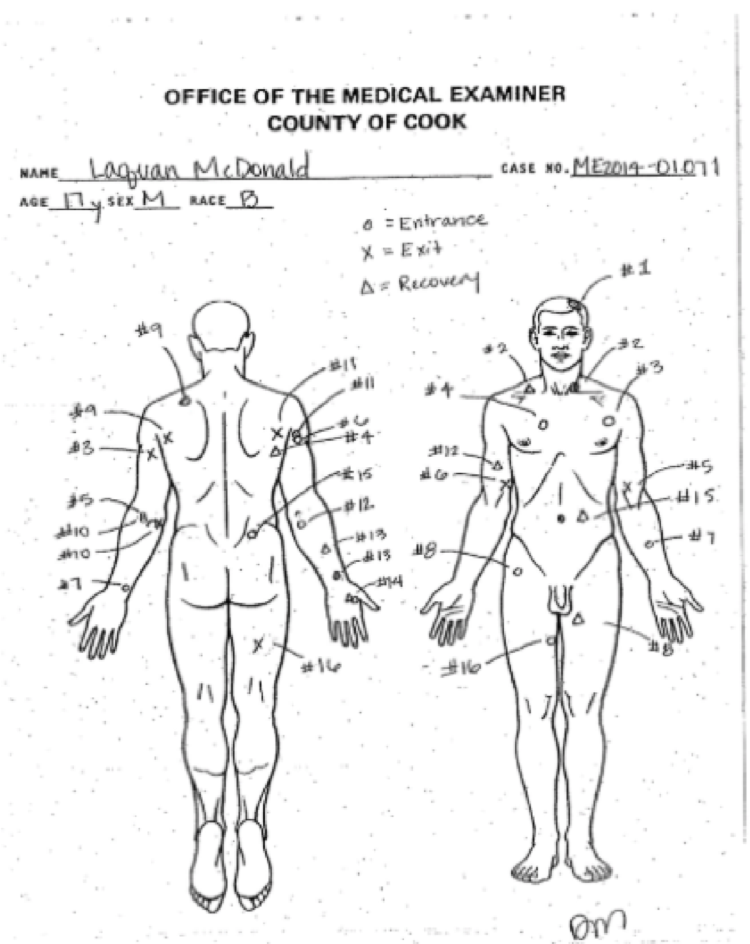
(Continued on page 2)

Notes: This figure presents an article from the FOP Lodge 7 newsletter dated September 2014. The article is from the President of the union and a follow-up to the article in August 2014. The article was front the page of the newsletter.

Figure 3.8: Newsletter September 2014

The New York Times

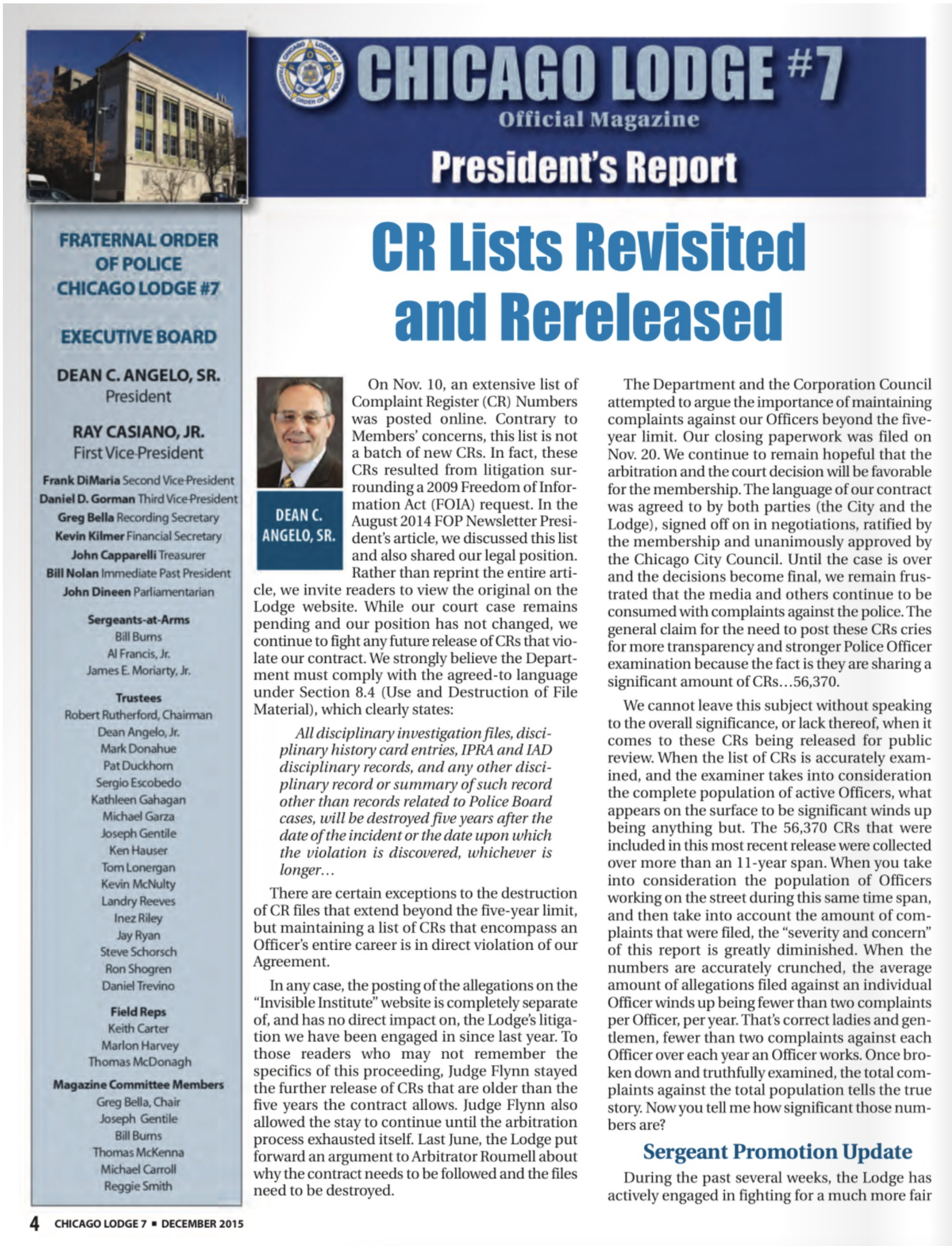
Video of Chicago Police Shooting a Teenager Is Ordered Released



An autopsy drawing of Laquan McDonald, 17, who was fatally shot by a Chicago police officer in October 2014. According to a report from the Cook County medical examiner's office, he was shot 16 times. Cook County Medical Examiner

Notes: This figure presents an article from the New York Times dated November 2015. The article is from the President of the union and a follow-up to the article in August 2014. The article covers the public release of the dash-cam footage of the murder of teenager Laquan McDonald by Jason Van Dyke, an officer from the Chicago Police Department.

Figure 3.9: New York Times - November 2015



CHICAGO LODGE #7
Official Magazine
President's Report

CR Lists Revisited and Rereleased

DEAN C. ANGELO, SR.

On Nov. 10, an extensive list of Complaint Register (CR) Numbers was posted online. Contrary to Members' concerns, this list is not a batch of new CRs. In fact, these CRs resulted from litigation surrounding a 2009 Freedom of Information Act (FOIA) request. In the August 2014 FOP Newsletter President's article, we discussed this list and also shared our legal position. Rather than reprint the entire article, we invite readers to view the original on the Lodge website. While our court case remains pending and our position has not changed, we continue to fight any future release of CRs that violate our contract. We strongly believe the Department must comply with the agreed-to language under Section 8.4 (Use and Destruction of File Material), which clearly states:

> *All disciplinary investigation files, disciplinary history card entries, IPRA and IAD disciplinary records, and any other disciplinary record or summary of such record other than records related to Police Board cases, will be destroyed five years after the date of the incident or the date upon which the violation is discovered, whichever is longer…*

There are certain exceptions to the destruction of CR files that extend beyond the five-year limit, but maintaining a list of CRs that encompass an Officer's entire career is in direct violation of our Agreement.

In any case, the posting of the allegations on the "Invisible Institute" website is completely separate of, and has no direct impact on, the Lodge's litigation we have been engaged in since last year. To those readers who may not remember the specifics of this proceeding, Judge Flynn stayed the further release of CRs that are older than the five years the contract allows. Judge Flynn also allowed the stay to continue until the arbitration process exhausted itself. Last June, the Lodge put forward an argument to Arbitrator Roumell about why the contract needs to be followed and the files need to be destroyed.

The Department and the Corporation Council attempted to argue the importance of maintaining complaints against our Officers beyond the five-year limit. Our closing paperwork was filed on Nov. 20. We continue to remain hopeful that the arbitration and the court decision will be favorable for the membership. The language of our contract was agreed to by both parties (the City and the Lodge), signed off on in negotiations, ratified by the membership and unanimously approved by the Chicago City Council. Until the case is over and the decisions become final, we remain frustrated that the media and others continue to be consumed with complaints against the police. The general claim for the need to post these CRs cries for more transparency and stronger Police Officer examination because the fact is they are sharing a significant amount of CRs…56,370.

We cannot leave this subject without speaking to the overall significance, or lack thereof, when it comes to these CRs being released for public review. When the list of CRs is accurately examined, and the examiner takes into consideration the complete population of active Officers, what appears on the surface to be significant winds up being anything but. The 56,370 CRs that were included in this most recent release were collected over more than an 11-year span. When you take into consideration the population of Officers working on the street during this same time span, and then take into account the amount of complaints that were filed, the "severity and concern" of this report is greatly diminished. When the numbers are accurately crunched, the average amount of allegations filed against an individual Officer winds up being fewer than two complaints per Officer, per year. That's correct ladies and gentlemen, fewer than two complaints against each Officer over each year an Officer works. Once broken down and truthfully examined, the total complaints against the total population tells the true story. Now you tell me how significant those numbers are?

Sergeant Promotion Update

During the past several weeks, the Lodge has actively engaged in fighting for a much more fair

Notes: This figure presents an article from the FOP Lodge 7 newsletter dated December 2015. The article is from the President of the union and a follow-up to the articles 2014 related to the released of the complaints.

Figure 3.10: Newsletter December 2015

CHICAGO LODGE #7

Official Magazine

President's Report

FRATERNAL ORDER OF POLICE CHICAGO LODGE #7

EXECUTIVE BOARD

Department of Justice

DEAN C. ANGELO, SR.

We have previously shared information relating to the Department of Justice (DOJ) and its examination of the CPD on the Lodge website. We also discussed it at the Dec. 15 General Membership meeting. What follows is our attempt to bring the entire membership as up to date as possible.

As many may know by now, on Dec. 15, the day following Illinois Attorney General Lisa Madigan's call for the DOJ's examination of the CPD, we were in conversation with National FOP President Chuck Canterbury to arrange a meeting with the DOJ. Prior to U.S. Attorney General Lynch's Dec. 7 announcement, we were 100-percent convinced that it was only a matter of time before the DOJ arrived in Chicago. Before the official announcement came out of Washington, and in order to be in the best position possible to protect our membership, we decided to be proactive and make every attempt to learn as much as possible about the DOJ's process. President Canterbury and National FOP Executive Director James Pasco successfully arranged a meeting on Dec. 9 with the DOJ.

We discussed several issues during the nearly hour-long meeting. We asked about the specifics of the DOJ's process as it relates to "pattern and practice" investigations. We also addressed concerns that several of our rank-and-file members have voiced. We inquired if the Lodge could somehow play an active role as the investigation progresses. We began our discussion with the overall climate, low morale and the everyday needs of our rank-and-file. We then expressed our frustrations with the recent finger pointing that had occurred in Chicago since the delayed release of the dash camera video.

We spoke about the political posturing that is regularly targeting the police officers who continue to perform with an amazing professionalism. We addressed the release of a felon the morning after striking a uniformed officer, the thousands of weapons recoveries that occur every year and a level of murders and gun violence not seen in any other city in America. We spoke about the vast amount of dash cameras that do not function as designed. We also shared that not only are there not enough Taser-certified officers (only one in five), but that there are also not enough Tasers to go around.

The DOJ could not (and would not) give a time frame for how long the investigation might take. Outside of Puerto Rico, Chicago is the largest department that the DOJ will examine. Initially, the DOJ will begin by collecting the General Orders concerning Use of Force, Disciplinary Investigations, IPRA's responsibilities, and, yes, the Contract. The DOJ will then begin interviewing civilians, CPD members and, at some point down the road, they will be involved in ride-a-longs with district personnel.

As the process of interviewing our members moves forward, the Lodge has offered to assist with the facilitation and has extended use of the Lodge building as a location to conduct some interviews. We realize that the comfort an officer would feel on Washington Boulevard is quite different compared to 35th and Michigan and/or the Federal Building. At the time of this writing, the DOJ was in favor of using Lodge 7 as a site for conducting interviews. As things continue to develop, we will continue to keep everyone informed.

To say that we owe our appreciation to President Canterbury and Director Pasco is a huge understatement. It is comforting for us to know that Lodge 7's relationship with the national office can provide our organization with this type of response and service.

Exhausted Accusatory Rhetoric

As our municipal, county and state politicians contemplate how to best go about designing a plan pertaining to the recent police-related incidents, they do so in a manner that further damages any remaining relationships with the Lodge 7 membership. They continue to distance themselves from us every time they get near a microphone or stage their next self-serving press conference.

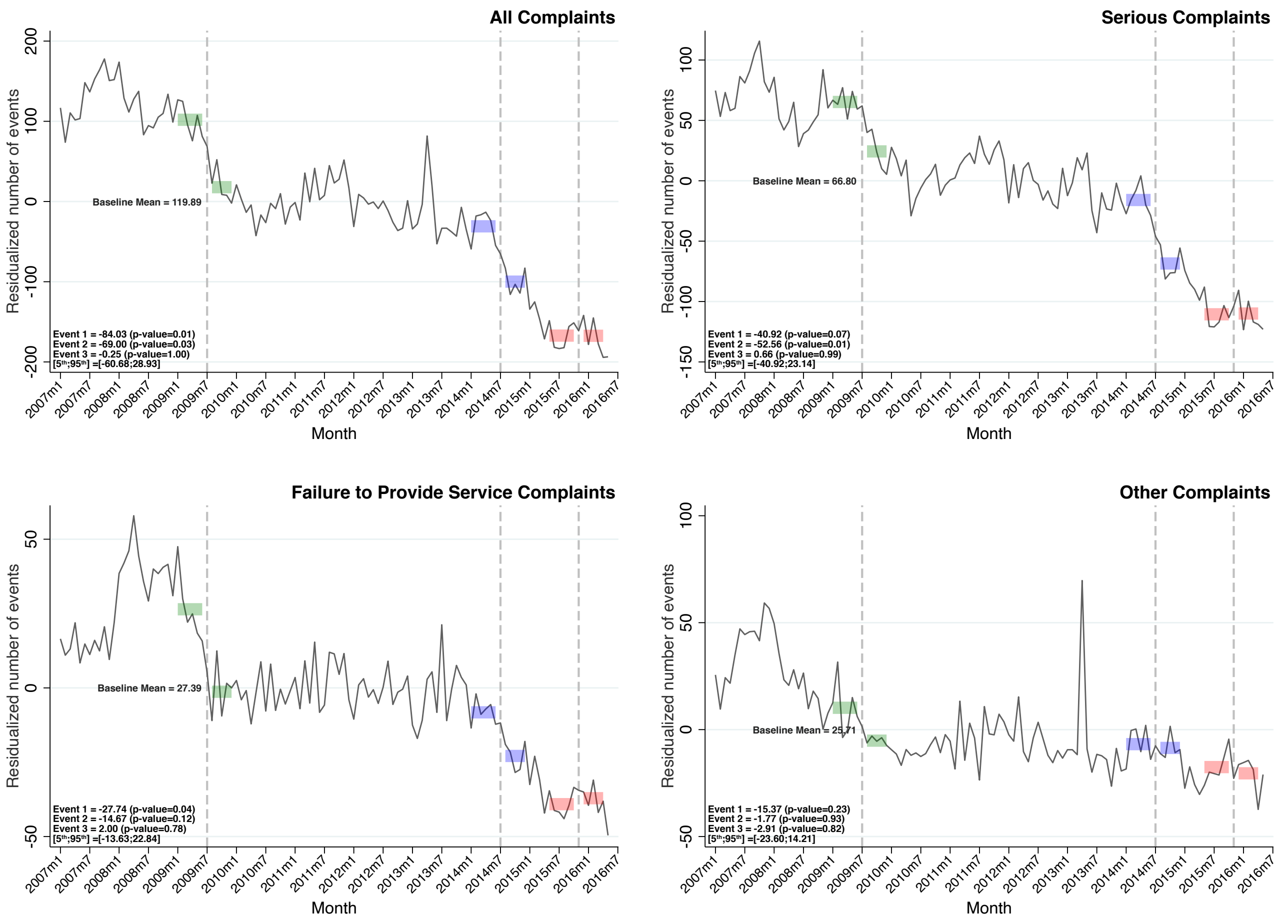
Many were quick to "Blame the Police, Demean the Police and Accuse the Police." As if it wasn't enough to blast the entire CPD, they then moved to "Blame the FOP, Demean the FOP and Accuse the FOP." Even now, it is apparent that it is easier for many of them to deflect their collective lack of familiarity with the contents of our Contract than to admit to their constituents that they voted on approving (our Contract received unanimous council approval) without ever reading. To accuse the Lodge of having membership protections that are undeserving, unnecessary and even illegal, further demonstrates that some have no knowledge of what is contained in our Agreement. It is typical for comments such as these to come from the uninformed.

It has become apparent that Lodge 7 is viewed as speaking on behalf of the collective body of Chicago Police Officers, no matter the rank. The Lodge must regularly address what is becoming a politically motivated attack on our membership. Recent

4 CHICAGO LODGE 7 ▪ JANUARY 2016

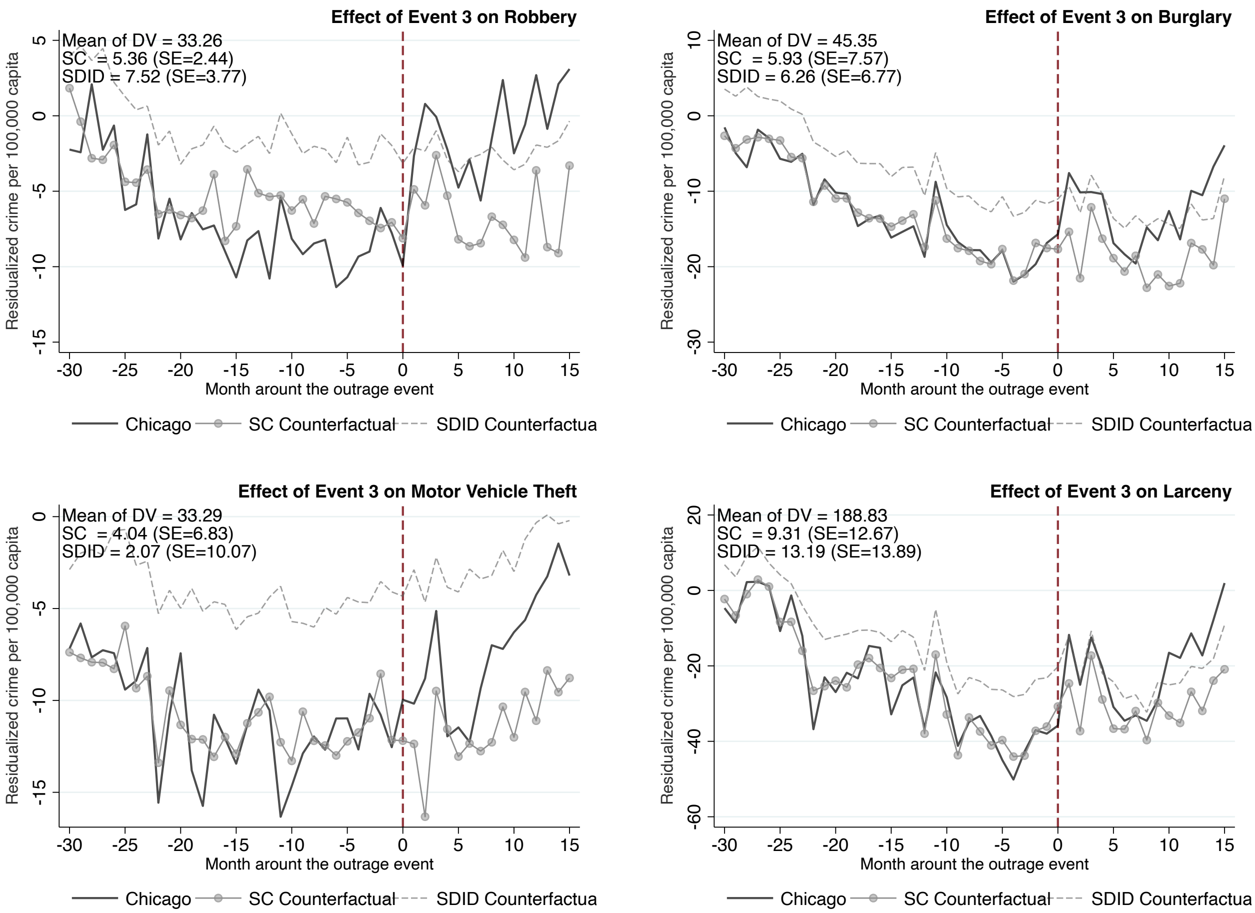
Notes: This figure presents an article from the FOP Lodge 7 newsletter dated January 2016. The articles is from the President of the union and discuss the DOJ investigation and the released of the dash-cam footage of the murder of teenager Laquan McDonald.

Figure 3.11: Newsletter January 2016



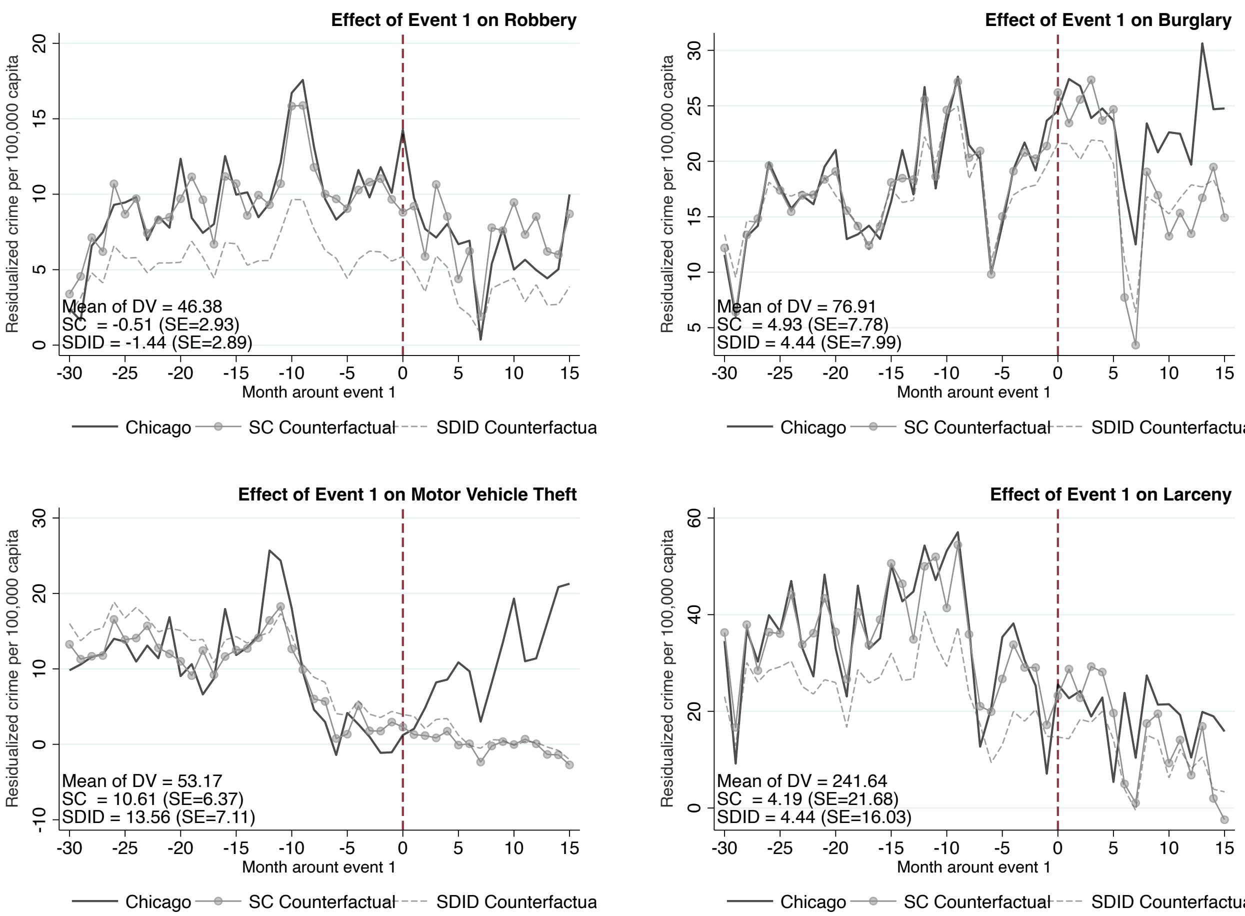
Notes: This figure presents the total number of misconduct allegations filed against a Chicago police officer from January 2007 to May 2016. The vertical dashed lines represent three events that have plausibly varied the level of oversight against the Chicago Police Department between 2007 and 2016. We form a test statistic for the impact of the events using the difference between the red horizontal bars. We omit months of the events of the interests because they show a mechanical change in the number of complaints the variation of oversight. We recompute the test statistic for every placebo date (N=70), where we define placebo windows as those with no event that changes the level of oversight. Finally, we report the 5th and 95th percentile values from the permutation tests and the p-values associated with each test. We account for seasonal patterns in outcomes by removing the month effect and reporting this regression's residuals. Finally, we report the mean of the dependent variable at baseline, i.e., before the first event. Section 3.4, in the main text, describes each event in detail.

Figure 3.12: Effect of Oversight and Outrage on Complaints



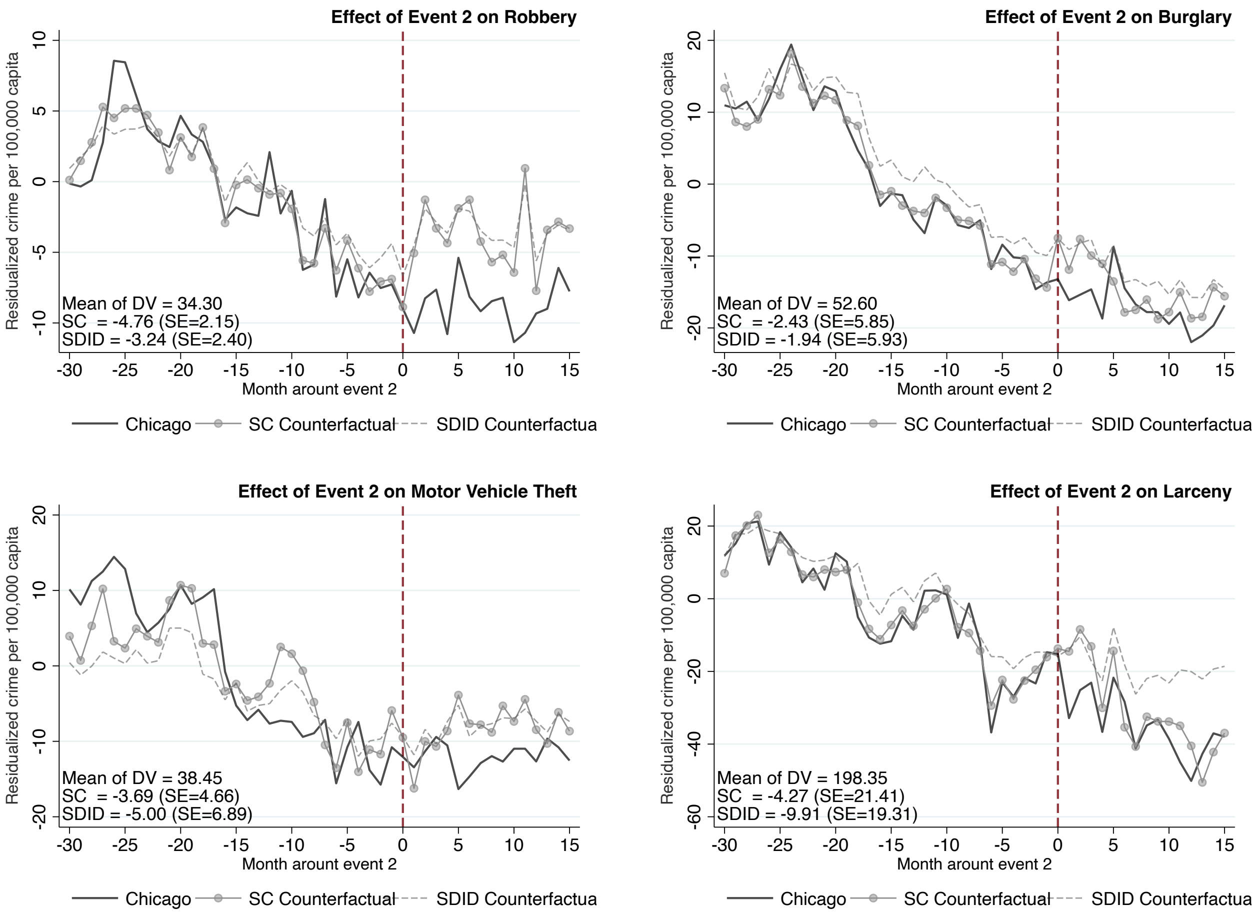
Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.13: Effect of Outrage on Different Crimes Category



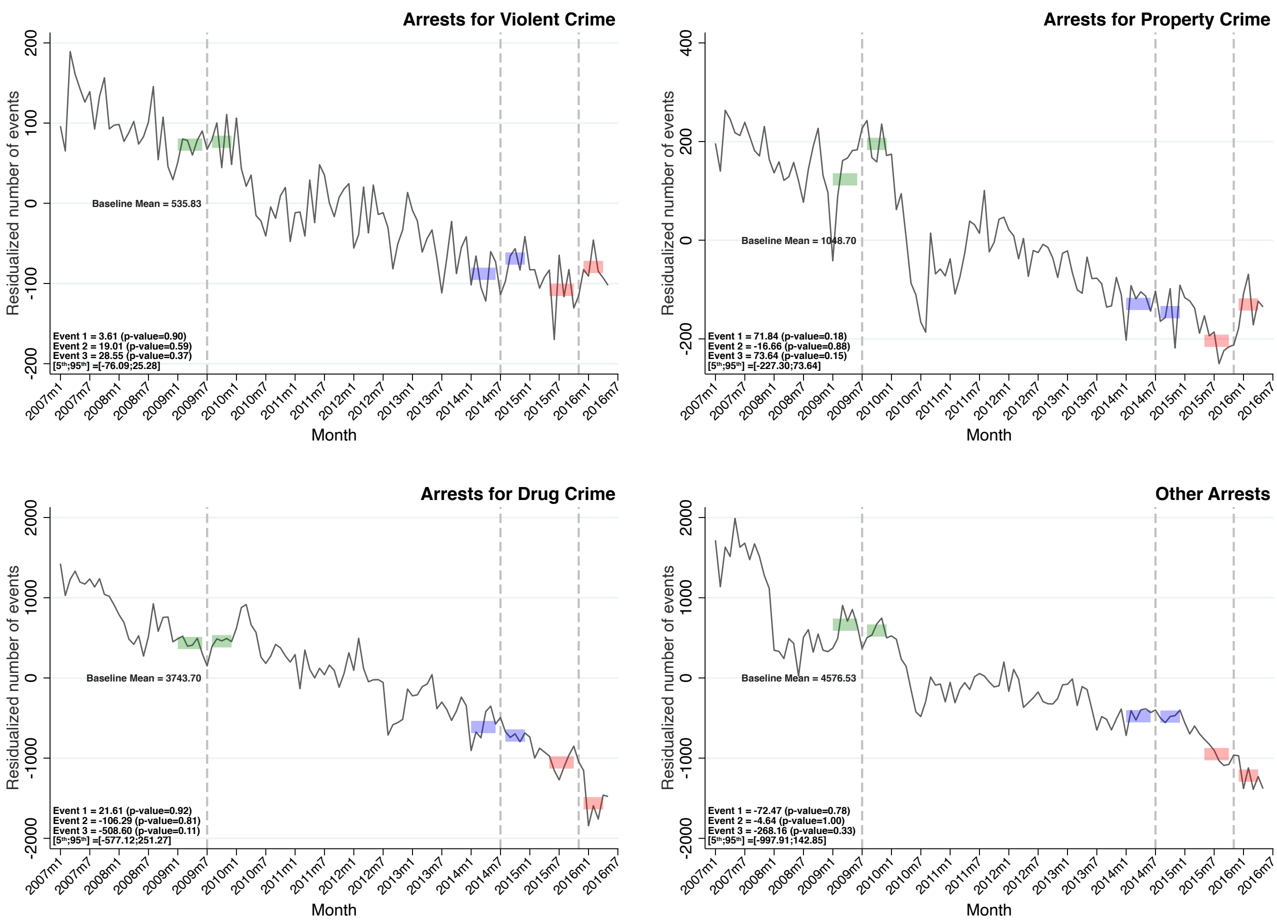
Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.14: Effect of Oversight Event 1 on Different Crimes Category



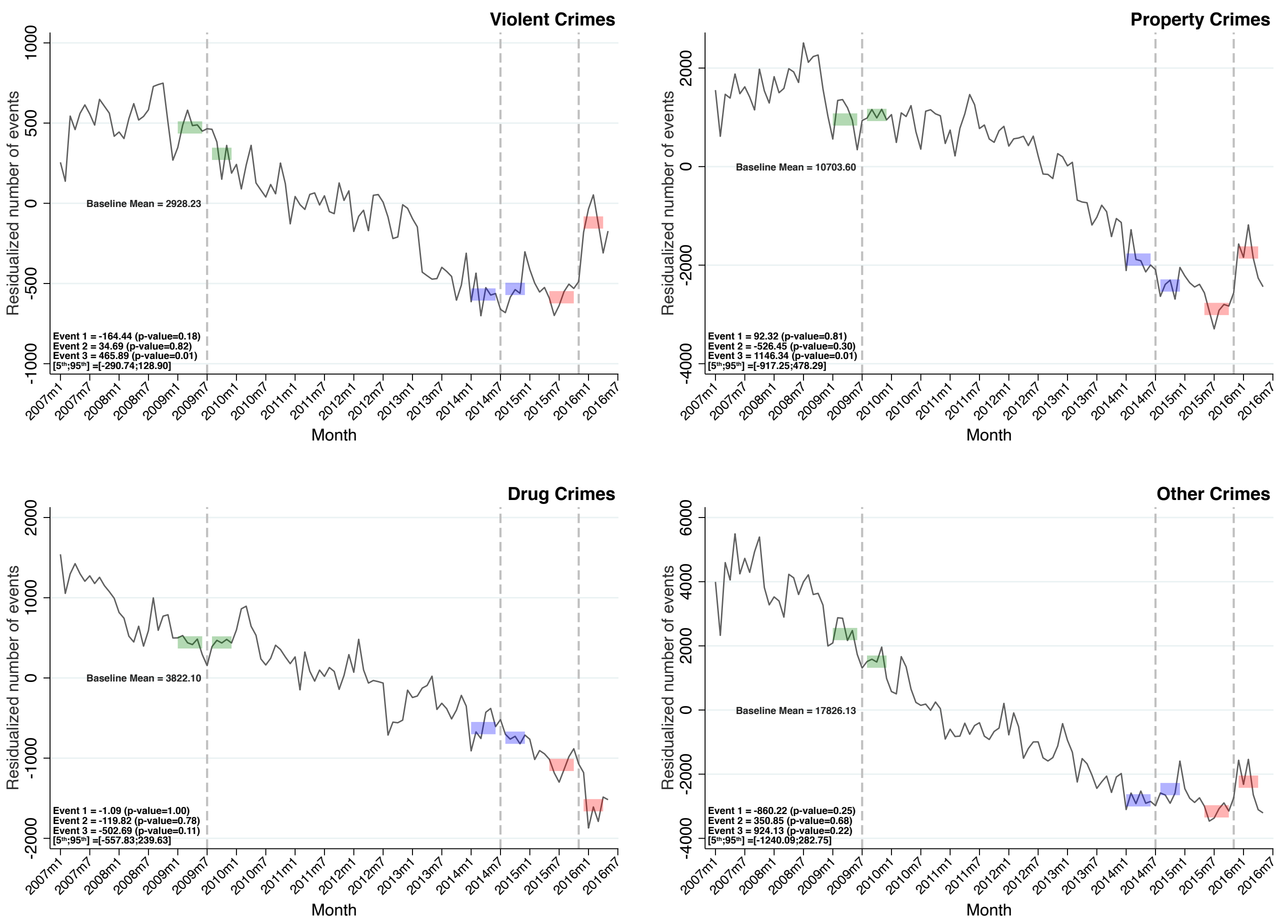
Notes: These figures present the crime trends per 100,000 population in Chicago and its synthetic difference-in-differences (SDID) and synthetic control (SC) counterfactuals. In addition, we account for seasonal patterns in crime outcomes, i.e., we remove the month effect and report the residuals of this regression (some outcomes can have negative values). The vertical dashed line represents the time of the event of interest. We report the mean of the dependent variable, the SDID and SC estimates. Standard errors are in parentheses.

Figure 3.15: Effect of Oversight Event 2 on Different Crimes Category



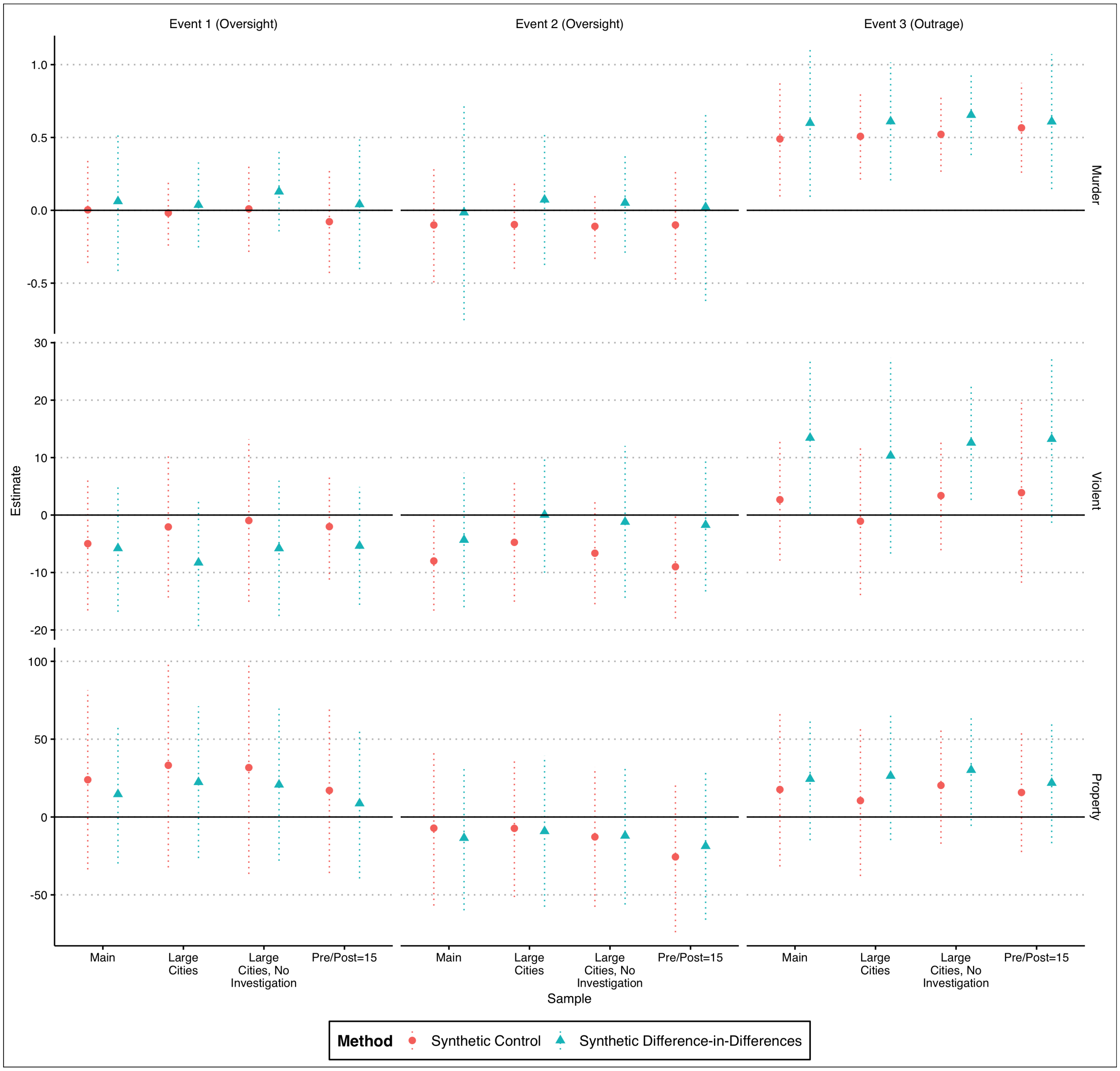
Notes: This figure presents the total number of arrests in Chicago from January 2007 to May 2016. The vertical dashed lines represent three events that have plausibly varied the level of oversight against the Chicago Police Department between 2007 and 2016. We form a test statistic for the impact of the events using the difference between the red horizontal bars. We omit months of the events of the interests because they show a mechanical change in the number of complaints the variation of oversight. We recompute the test statistic for every placebo date (N=70), where we define placebo windows as those with no event that changes the level of oversight. Finally, we report the 5th and 95th percentile values from the permutation tests and the p-values associated with each test. We account for seasonal patterns in outcomes by removing the month effect and reporting this regression's residuals. Finally, we report the mean of the dependent variable at baseline, i.e., before the first event. Section 3.4, in the main text, describes each event in detail.

Figure 3.16: Effect of Oversight and Outrage on Arrests



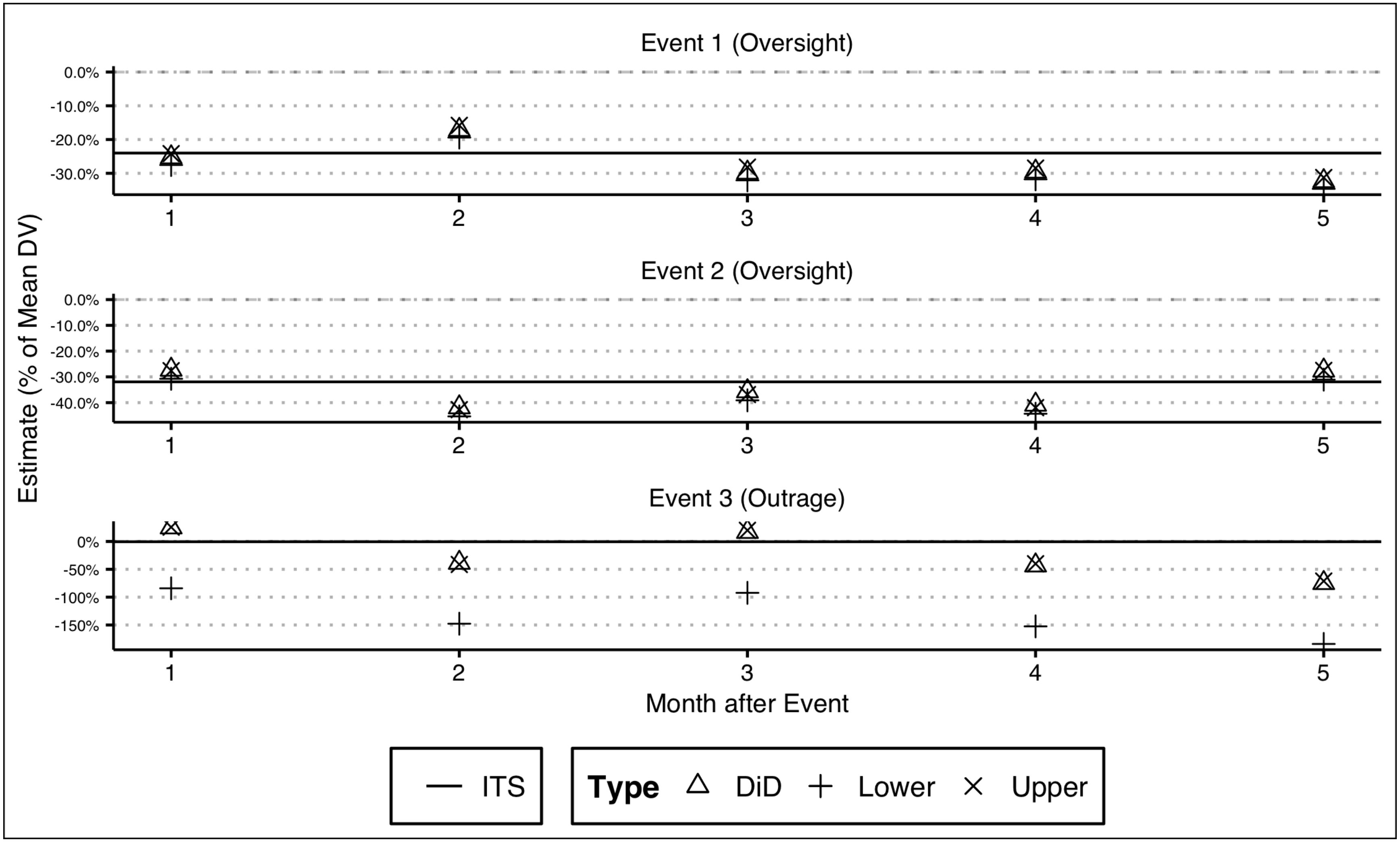
Notes: This figure presents the total number of crimes in Chicago from January 2007 to May 2016. The vertical dashed lines represent three events that have plausibly varied the level of oversight against the Chicago Police Department between 2007 and 2016. We form a test statistic for the impact of the events using the difference between the red horizontal bars. We omit months of the events of the interests because they show a mechanical change in the number of complaints the variation of oversight. We recompute the test statistic for every placebo date (N=70), where we define placebo windows as those with no event that changes the level of oversight. Finally, we report the 5th and 95th percentile values from the permutation tests and the p-values associated with each test. We account for seasonal patterns in outcomes by removing the month effect and reporting this regression's residuals. Finally, we report the mean of the dependent variable at baseline, i.e., before the first event. Section 3.4, in the main text, describes each event in detail.

Figure 3.17: Effect of Oversight and Outrage on Crimes



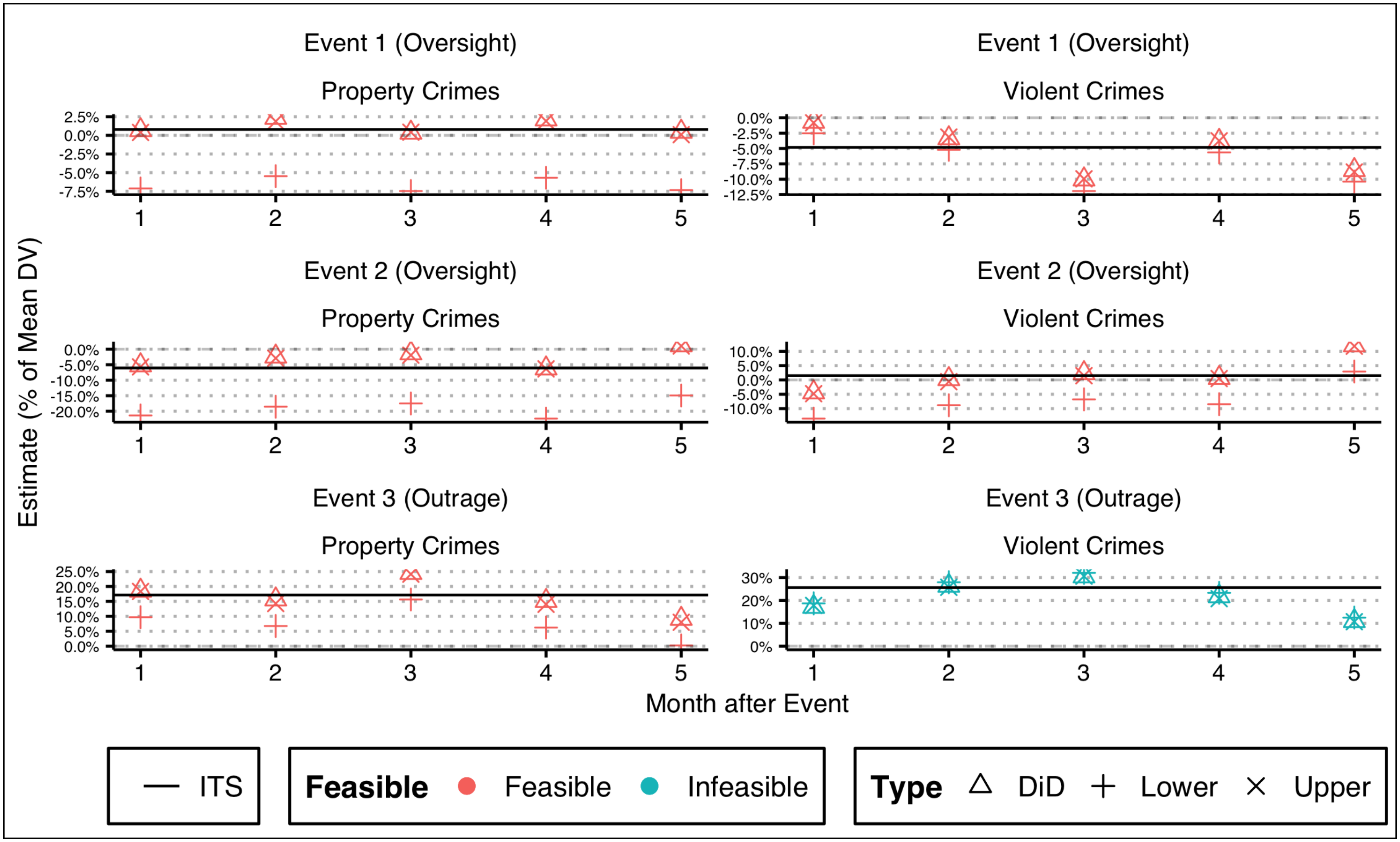
Notes: This figure displays the results of performing synthetic difference-in-differences (circle) and synthetic control (triangle) estimations. We report the 95% confidence intervals. Each set of results refers to an event, with four different samples/specifications notes on each X-axis, including the main results. Large cities refer to restricting the sample of control cities to those with at least 250,000 population; Large cities are cities with at least 500,000 population; no investigations further restrict the large cities sample to those without DOJ investigation. The Pre/Post=15 refers to using 15 pre-periods (30 in the main specification) and 15 post periods. The main estimation includes 30 pre-intervention periods and six post-intervention periods. Each Y-axis refers to a different crime type (murder, violent index, and property index crime).

Figure 3.18: National Sample Robustness



Notes: This figure displays the results of performing the [227] difference-in-differences estimation and bounding exercise on the effect of each event on the number of complaints. Each plot contains the estimated lower bound (cross), upper bound (X), and difference-in-differences estimate (triangle) for each month after the respective event. For reference, the ITS estimate from the main results (Table 3.2) is included as a flat line. The Y-axes are scaled to the percent change relative to the mean dependent variable. All the bounds are feasible.

Figure 3.19: Manski and Pepper (2018) Bounds for Effect on All Complaints



Notes: This figure displays the results of performing the [227] difference-in-differences estimation and bounding exercise on the effect of each event (rows) and property and violent crime (columns). Each plot contains the estimated lower bound (cross), upper bound (X), and difference-in-differences estimate (triangle) for each month after the respective event. For reference, the ITS estimate from the main results (Table 3.2) is included as a flat line. Estimates are colored by whether they are feasible, meaning the lower bound is smaller than the upper bound. The Y-axes are scaled to the percent change relative to the mean dependent variable.

Figure 3.20: Manski and Pepper (2018) Bounds for Effect on Crime

	$\tau = 4$	$\tau = 8$	$\tau = 4$	$\tau = 8$	$\tau = 4$	$\tau = 8$
	Effect of Event (% Change Relative to Mean)					
Outcome	Oversight (1)	Oversight (1)	Oversight (2)	Oversight (2)	Outrage (3)	Outrage (3)
Complaints						
All	-0.208 (p = 0.03)	-0.257 (p = 0.01)	-0.318 (p = 0.01)	-0.346 (p = 0.06)	0.007 (p = 1)	-0.024 (p = 0.93)
Serious	-0.205 (p = 0.05)	-0.226 (p = 0.04)	-0.454 (p = 0.01)	-0.487 (p = 0.01)	0.13 (p = 0.84)	-0.104 (p = 0.88)
FPS	-0.281 (p = 0.06)	-0.38 (p = 0.04)	-0.301 (p = 0.13)	-0.385 (p = 0.12)	0.126 (p = 0.71)	0.068 (p = 0.82)
Other	-0.136 (p = 0.34)	-0.204 (p = 0.25)	-0.052 (p = 0.84)	0.001 (p = 1)	-0.111 (p = 0.59)	-0.006 (p = 0.99)
Public						
Tribune Articles Related to CPD	-0.613 (p = 0.52)	-0.109 (p = 0.88)	-0.074 (p = 0.99)	0.451 (p = 0.66)	2.602 (p = 0.01)	3.274 (p = 0.01)
Crime						
Violent	-0.065 (p = 0.11)	-0.052 (p = 0.28)	0.03 (p = 0.66)	-0.003 (p = 1)	0.196 (p = 0.01)	0.221 (p = 0.01)
Property	0.009 (p = 0.82)	0.001 (p = 0.99)	-0.06 (p = 0.44)	-0.102 (p = 0.18)	0.162 (p = 0.01)	0.146 (p = 0.01)
Arrests/Stops/Force						
Violent	0 (p = 1)	0.026 (p = 0.76)	0.057 (p = 0.52)	0.027 (p = 0.87)	0.05 (p = 0.57)	0.072 (p = 0.43)
Property	0.021 (p = 0.76)	0.076 (p = 0.21)	-0.034 (p = 0.7)	-0.035 (p = 0.84)	0.119 (p = 0.11)	0.087 (p = 0.31)
Drug	0.017 (p = 0.87)	-0.006 (p = 0.97)	-0.079 (p = 0.53)	-0.079 (p = 0.54)	-0.253 (p = 0.08)	-0.272 (p = 0.09)
Other	-0.04 (p = 0.43)	0.002 (p = 0.99)	-0.014 (p = 0.85)	-0.004 (p = 0.97)	-0.065 (p = 0.42)	-0.113 (p = 0.3)
Stops			-0.171 (p = 0.35)	-0.085 (p = 0.57)	-0.722 (p = 0.02)	-0.782 (p = 0.02)
Force	0.021 (p = 0.63)	-0.002 (p = 0.97)	-0.023 (p = 0.58)	0.029 (p = 0.78)	-0.136 (p = 0.03)	-0.127 (p = 0.03)

Table 3.5: Effect of Oversight and Outrage on Outcomes

Note: Table contains the ITS results from estimating equation 3.1 with 4 (8) pre-event months and 3 (7) post-event months for each outcome in the odd (even) columns. Each set of two columns refers to an event. Stop data could not be obtained pre-2010, so the effect of Event 1 on stops could not be estimated. Coefficients are all scaled as a share relative to the mean of the dependent variable in the 6-months before the event. Dependent variables are residualized by month to remove seasonality. P-values are computed by ranking the effects the 3 events along with the effects of 76 and 64 placebo periods, which exclude the period of the events and the events themselves, for $\tau = 4$ and $\tau = 8$ respectively.

References

[1] W. Sawyer and P. Wagner, “Mass incarceration: The whole pie 2022,” Prison Policy Initiative, Press Release, Mar. 2022.

[2] P. Wagner and B. Rabuy, “Following the money of mass incarceration,” Prison Policy Initiative, Tech. Rep., Jan. 2017.

[3] Pew, *Use of electronic offender-tracking devices expands sharply*, http://pew.org/2cpDaNx, Sep. 2016.

[4] K. Weisburd *et al.*, “Electronic prisons: The operation of ankle monitoring in the criminal legal system,” The George Washington Law School, Report, Oct. 2021, p. 54.

[5] M. Guevara, “Miami-dade corrections and rehabilitation department, biannual report 2013 and 2014,” Miami-Dade Corrections and Rehabilitation Department, Biannual Report, 2014, p. 32.

[6] E. Hager, *Where coronavirus is surging - and electronic surveillance, too*, https://www.themarshallproject. coronavirus-is-surging-and-electronic-surveillance-too, Nov. 2020.

[7] M. Barajas, *As covid-19 cases surge, some incarcerated people remain behind bars after making bail*, https://www.texasobserver.org/pretrial-covid-dallas-electronic-monitoring/, Jun. 2020.

[8] T. C. Federation, *Cook county seeks consulting services to review electronic monitoring practices*, https://www.civicfed.org/civic-federation/blog/cook-county-seeks-consulting-services-review-electronic-monitoring-practices, May 2020.

[9] K. Weisburd and A. Virani, *Op-ed: The monster of incarceration quietly expands through ankle monitors*, https://www.latimes.com/opinion/story/2022-03-15/electronic-monitoring-pretrial-bail-mass-incarceration, Mar. 2022.

[10] A. Virani, “Pretrial electronic monitoring in los angeles county,” UCLA School of Law, Los Angeles, Report, 2022.

[11] APPA, *Appa response to criticisms of electronic monitoring*, Mar. 2020.

[12] J. Kilgore, “Progress or more of the same? electronic monitoring and parole in the age of mass incarceration,” *Critical Criminology*, vol. 21, no. 1, pp. 123–139, Mar. 2013.

[13] M. DeMichele, “Electronic monitoring: It is a tool, not a silver bullet electronic monitoring on social welfare dependence: Policy essay,” *Criminology and Public Policy*, vol. 13, no. 3, pp. 393–400, 2014.

[14] PRI, “Ten-point plan on reducing pre-trial detention,” Penal Reform International, Tech. Rep., 2016.

[15] M. Alexander, *The newest jim crow*, https://www.nytimes.com/2018/11/08/opinion/sunday/criminal-justice-reforms-race-technology.html, Nov. 2018.

[16] R. Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK ; Medford, MA: Polity, 2019, ISBN: 978-1-5095-2643-7.

[17] P. R. Lockhart, *Thousands of americans are jailed before trial. a new report shows the lasting impact.* https://www.vox.com/2019/5/7/18527237/pretrial-detention-jail-bail-reform-vera-institute-report, May 2019.

[18] K. Weisburd, “Punitive surveillance,” *Virginia Law Review*, vol. 108, p. 75, 2022.

[19] T. R. Miller, M. A. Cohen, D. I. Swedler, B. Ali, and D. V. Hendrie, “Incidence and costs of personal and property crimes in the usa, 2017,” *Journal of Benefit-Cost Analysis*, vol. 12, no. 1, pp. 24–54, 2021.

[20] R. Di Tella and E. Schargrodsky, “Criminal recidivism after prison and electronic monitoring,” *Journal of Political Economy*, vol. 121, no. 1, pp. 28–73, Feb. 2013.

[21] A. Henneguelle, B. Monnery, and A. Kensey, “Better at home than in prison? the effects of electronic monitoring on recidivism in france,” *The Journal of Law and Economics*, vol. 59, no. 3, pp. 629–667, Aug. 2016.

[22] J. Williams and D. Weatherburn, “Can electronic monitoring reduce reoffending?” *The Review of Economics and Statistics*, vol. 104, no. 2, pp. 232–245, Mar. 2022.

[23] O. Marie, “Early release from prison and recidivism: A regression discontinuity approach,” Working Paper, Jul. 2008, p. 22.

[24] A. Ouss, “Sensitivity analyses in economics of crime: Do monitored suspended sentences reduce recidivism?” Working Paper, Jul. 2013, p. 26.

[25] L. H. Andersen and S. H. Andersen, “Effect of electronic monitoring on social welfare dependence,” *Criminology & Public Policy*, vol. 13, no. 3, pp. 349–379, 2014.

[26] J. Grenet, H. Grönqvist, and S. Niknami, “The effects of electronic monitoring on offenders and their families,” Working Paper, Aug. 2022, p. 40.

[27] J. Belur, A. Thornton, L. Tompson, M. Manning, A. Sidebottom, and K. Bowers, "A systematic review of the effectiveness of the electronic monitoring of offenders," *Journal of Criminal Justice*, vol. 68, p. 101 686, May 2020.

[28] A. Gupta, C. Hansman, and E. Frenchman, "The heavy costs of high bail: Evidence from judge randomization," *Journal of Legal Studies*, vol. 45, no. 2, pp. 471–505, Jun. 2016.

[29] E. Leslie and N. G. Pope, "The unintended impact of pretrial detention on case outcomes: Evidence from new york city arraignments," *The Journal of Law and Economics*, vol. 60, no. 3, pp. 529–557, Aug. 2017.

[30] W. Dobbie, J. Goldin, and C. S. Yang, "The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges," *American Economic Review*, vol. 108, no. 2, pp. 201–240, Feb. 2018.

[31] M. Stevenson, "Distortion of justice: How the inability to pay bail affects case outcomes," *The Journal of Law, Economics, and Organization*, vol. 34, no. 4, pp. 511–542, Nov. 2018.

[32] D. Arnold, W. Dobbie, and C. S. Yang, "Racial bias in bail decisions," *The Quarterly Journal of Economics*, vol. 133, no. 4, pp. 1885–1932, Nov. 2018.

[33] A. Y. Agan, J. L. Doleac, and A. Harvey, "Misdemeanor prosecution," National Bureau of Economic Research, Tech. Rep. w28600, Mar. 2021.

[34] D. Arnold, W. Dobbie, and P. Hull, "Measuring racial discrimination in bail decisions," *American Economic Review*, vol. 112, no. 9, pp. 2992–3038, Sep. 2022.

[35] J. R. Kling, "Incarceration length, employment, and earnings," *American Economic Review*, vol. 96, no. 3, pp. 863–876, Jun. 2006.

[36] J. J. Doyle Jr., "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care," *Journal of Political Economy*, vol. 116, no. 4, pp. 746–770, Aug. 2008.

[37] A. Aizer and J. J. Doyle, "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," *The Quarterly Journal of Economics*, vol. 130, no. 2, pp. 759–803, May 2015.

[38] M. Mueller-Smith, "The criminal and labor market impacts of incarceration," Working Paper, Aug. 2015, p. 59.

[39] W. Dobbie and J. Song, "Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection," *American Economic Review*, vol. 105, no. 3, pp. 1272–1311, Mar. 2015.

[40] B. Frandsen, L. Lefgren, and E. Leslie, “Judging judge fixed effects,” National Bureau of Economic Research, Cambridge, MA, Tech. Rep. w25528, Feb. 2019.

[41] M. Bhuller, G. B. Dahl, K. V. L, and M. Mogstad, “Incarceration, recidivism, and employment,” *Journal of Political Economy*, vol. 128, no. 4, pp. 1269–1324, Apr. 2020.

[42] S. Norris, M. Pecenco, and J. Weaver, “The effects of parental and sibling incarceration: Evidence from ohio,” *American Economic Review*, vol. 111, no. 9, pp. 2926–2963, Sep. 2021.

[43] C. Arteaga, “Parental incarceration and children’s educational attainment,” Working Paper, Apr. 2021, p. 68.

[44] A. Jordan, E. Karger, and D. A. Neal, *Heterogeneous impacts of sentencing decisions*, SSRN Scholarly Paper, Rochester, NY, Aug. 2022.

[45] F. Goncalves and S. Mello, “Should the punishment fit the crime? deterrence and retribution in law enforcement,” Working Paper, Mar. 2022, p. 72.

[46] M. Gross and E. J. Baron, “Temporary stays and persistent gains: The causal effects of foster care,” *American Economic Journal: Applied Economics*, vol. 14, no. 2, pp. 170–199, Apr. 2022.

[47] W. Dobbie and C. S. Yang, “The us pretrial system: Balancing individual rights and public interests,” *Journal of Economic Perspectives*, vol. 35, no. 4, pp. 49–70, Nov. 2021.

[48] S. L. Myers, “The economics of bail jumping,” *The Journal of Legal Studies*, vol. 10, no. 2, pp. 381–396, 1981.

[49] A. Hawken and M. Kleiman, *Managing drug involved probationers with swift and certain sanctions: Evaluating hawaii’s hope: (513502010-001)*, 2009.

[50] B. Kilmer, N. Nicosia, P. Heaton, and G. Midgette, “Efficacy of frequent monitoring with swift, certain, and modest sanctions for violations: Insights from south dakota’s 24/7 sobriety project,” *American Journal of Public Health*, vol. 103, no. 1, Jan. 2013.

[51] I. Kuziemko, “How should inmates be released from prison? an assessment of parole versus fixed-sentence regimes *,” *The Quarterly Journal of Economics*, vol. 128, no. 1, pp. 371–424, Feb. 2013.

[52] E. K. Rose, “Who gets a second chance? effectiveness and equity in supervision of criminal offenders*,” *The Quarterly Journal of Economics*, vol. 136, no. 2, pp. 1199–1253, May 2021.

[53] M. LaForest, "Early release and parole conditions at the margin," Working Paper, Mar. 2021.

[54] W. Arbour and S. Marchand, "Parole, recidivism, and the role of supervised transition," Working Paper, May 2022, p. 38.

[55] A. A. Alchian and H. Demsetz, "Production, information costs, and economic organization," *The American Economic Review*, vol. 62, no. 5, pp. 777–795, 1972.

[56] B. Holmstrom and P. Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *Journal of Law, Economics, & Organization*, vol. 7, pp. 24–52, 1991.

[57] G. S. Becker and G. J. Stigler, "Law enforcement, malfeasance, and compensation of enforcers," *The Journal of Legal Studies*, vol. 3, no. 1, pp. 1–18, 1974.

[58] J. A. Mirrlees, "The optimal structure of incentives and authority within an organization," *The Bell Journal of Economics*, vol. 7, no. 1, pp. 105–131, 1976.

[59] B. Holmström, "Moral hazard and observability," *The Bell Journal of Economics*, vol. 10, no. 1, pp. 74–91, 1979.

[60] S. J. Grossman and O. D. Hart, "An analysis of the principal-agent problem," *Econometrica*, vol. 51, no. 1, pp. 7–45, 1983.

[61] C. Shapiro and J. E. Stiglitz, "Equilibrium unemployment as a worker discipline device," *The American Economic Review*, vol. 74, no. 3, pp. 433–444, 1984.

[62] C. Prendergast, "The provision of incentives in firms," *Journal of Economic Literature*, vol. 37, no. 1, pp. 7–63, Mar. 1999.

[63] C. Lum and D. S. Nagin, "Reinventing american policing," *Crime and Justice*, vol. 46, no. 1, pp. 339–393, Jan. 2017.

[64] M. A. R. Kleiman, *When Brute Force Fails: How to Have Less Crime and Less Punishment*. Princeton University Press, 2009, ISBN: 978-0-691-14864-9.

[65] S. N. Durlauf and D. S. Nagin, "Imprisonment and crime," *Criminology & Public Policy*, vol. 10, no. 1, pp. 13–54, 2011.

[66] N. Persico, "Racial profiling, fairness, and effectiveness of policing," *American Economic Review*, vol. 92, no. 5, pp. 1472–1497, Dec. 2002.

[67] C. F. Manski, "Optimal search profiling with linear deterrence," *The American Economic Review*, vol. 95, no. 2, pp. 122–126, 2005.

[68] J. Dominitz and J. Knowles, "Crime minimisation and racial bias: What can we learn from police search data?" *The Economic Journal*, vol. 116, no. 515, Nov. 2006.

[69] D. Bjerk, "Racial profiling, statistical discrimination, and the effect of a colorblind policy on the crime rate," *Journal of Public Economic Theory*, vol. 9, no. 3, pp. 521–545, 2007.

[70] B. Feigenberg and C. Miller, "Would eliminating racial disparities in motor vehicle searches have efficiency costs?*," *The Quarterly Journal of Economics*, vol. 137, no. 1, pp. 49–113, Feb. 2022.

[71] G. S. Becker, "Crime and punishment: An economic approach," *Journal of Political Economy*, vol. 76, no. 2, pp. 169–217, 1968.

[72] P. J. Cook, "The clearance rate as a measure of criminal justice system effectiveness," *Journal of Public Economics*, vol. 11, no. 1, pp. 135–142, Feb. 1979.

[73] I. Ehrlich, "Crime, punishment, and the market for offenses," *Journal of Economic Perspectives*, vol. 10, no. 1, pp. 43–67, Feb. 1996.

[74] J. J. Donohue, "Economic models of crime and punishment," *Social Research*, vol. 74, no. 2, pp. 379–412, 2007.

[75] D. S. Nagin, "Deterrence: A review of the evidence by a criminologist for economists," *Annual Review of Economics*, vol. 5, no. 1, pp. 83–105, Aug. 2013.

[76] A. Chalfin and J. McCrary, "Criminal deterrence: A review of the literature," *Journal of Economic Literature*, vol. 55, no. 1, pp. 5–48, Mar. 2017.

[77] M. C. Mungan, "The certainty versus the severity of punishment, repeat offenders, and stigmatization," *Economics Letters*, vol. 150, pp. 126–129, Jan. 2017.

[78] C. Lum *et al.*, "Body-worn cameras' effects on police officers and citizen behavior: A systematic review," *Campbell Systematic Reviews*, vol. 16, no. 3, 2020.

[79] N. Jensen, E. Lyons, E. Chebelyon, and R. L. Bras, "Conspicuous monitoring and remote work," *Journal of Economic Behavior & Organization*, vol. 176, pp. 489–511, Aug. 2020.

[80] J. Tirole, "Digital dystopia," *American Economic Review*, vol. 111, no. 6, pp. 2007–2048, Jun. 2021.

[81] M. Beraja, D. Y. Yang, and N. Yuchtman, "Data-intensive innovation and the state: Evidence from ai firms in china," National Bureau of Economic Research, Working Paper 27723, Aug. 2020.

[82] M. Beraja, A. Kao, D. Y. Yang, and N. Yuchtman, “Ai-tocracy,” National Bureau of Economic Research, Working Paper 29466, Nov. 2021.

[83] D. Acemoglu, *Harms of ai*, Working Paper, Sep. 2021.

[84] M. Barbaro *et al.*, “The rise of workplace surveillance,” *The New York Times*, Aug. 2022.

[85] L. J. Kirkeboen, E. Leuven, and M. Mogstad, “Field of study, earnings, and self-selection,” *The Quarterly Journal of Economics*, vol. 131, no. 3, pp. 1057–1111, Aug. 2016.

[86] P. Kline and C. R. Walters, “Evaluating public programs with close substitutes: The case of head start,” *The Quarterly Journal of Economics*, vol. 131, no. 4, pp. 1795–1848, Nov. 2016.

[87] J. Mountjoy, “Community colleges and upward mobility,” *American Economic Review*, vol. 112, no. 8, pp. 2580–2630, Aug. 2022.

[88] A. Björklund and R. Moffitt, “The estimation of wage gains and welfare gains in self-selection models,” *The Review of Economics and Statistics*, vol. 69, no. 1, pp. 42–49, 1987.

[89] J. Heckman and E. Vytlacil, “Local instrumental variables and latent variable models for identifying and bounding treatment effects,” *Proceedings of the National Academy of Sciences*, vol. 96, no. 8, pp. 4730–4734, Apr. 1999.

[90] G. B. Dahl, “Mobility and the return to education: Testing a roy model with multiple markets,” *Econometrica*, vol. 70, no. 6, pp. 2367–2420, 2002.

[91] J. Heckman and E. Vytlacil, “Structural equations, treatment effects, and econometric policy evaluation,” *Econometrica*, vol. 73, no. 3, pp. 669–738, 2005.

[92] J. Heckman, S. Urzua, and E. Vytlacil, “Instrumental variables in models with multiple outcomes: The general unordered case,” *Annales d’Économie et de Statistique*, no. 91/92, pp. 151–174, 2008.

[93] Moffitt, “Estimating marginal treatment effects in heterogeneous populations,” *Annales d’Économie et de Statistique*, no. 91/92, p. 239, 2008.

[94] P. Carneiro and S. Lee, “Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality,” *Journal of Econometrics*, vol. 149, no. 2, pp. 191–208, Apr. 2009.

[95] C. N. Brinch, M. Mogstad, and M. Wiswall, “Beyond late with a discrete instrument,” *Journal of Political Economy*, vol. 125, no. 4, pp. 985–1039, Aug. 2017.

[96] S. Lee and B. Salanié, “Identifying effects of multivalued treatments,” *Econometrica*, vol. 86, no. 6, pp. 1939–1963, 2018.

[97] M. Mogstad and A. Torgovitsky, “Identification and extrapolation of causal effects with instrumental variables,” *Annual Review of Economics*, vol. 10, no. 1, pp. 577–613, Aug. 2018.

[98] M. Bhuller and H. Sigstad, *2sls with multiple treatments*, May 2022. arXiv: `2205.07836 [econ]`.

[99] T. Cornelissen, C. Dustmann, A. Raute, and U. Schönberg, “From late to mte: Alternative methods for the evaluation of policy interventions,” *Labour Economics*, SOLE/EALE Conference Issue 2015, vol. 41, pp. 47–60, Aug. 2016.

[100] M. E. Andresen, “Exploring marginal treatment effects: Flexible estimation using stata,” *The Stata Journal: Promoting communications on statistics and Stata*, vol. 18, no. 1, pp. 118–158, Mar. 2018.

[101] E. K. Rose and Y. Shem-Tov, “How does incarceration affect reoffending? estimating the dose-response function,” *Journal of Political Economy*, Jul. 2021.

[102] J. Heckman, S. Urzua, and E. Vytlacil, “Understanding instrumental variables in models with essential heterogeneity,” *The Review of Economics and Statistics*, vol. 88, no. 3, pp. 389–432, 2006.

[103] M. Mogstad, A. Torgovitsky, and C. R. Walters, “The causal interpretation of two-stage least squares with multiple instrumental variables,” *American Economic Review*, vol. 111, no. 11, pp. 3663–3698, Nov. 2021.

[104] T. Cornelissen, C. Dustmann, A. Raute, and U. Schönberg, “Who benefits from universal child care? estimating marginal returns to early child care attendance,” *Journal of Political Economy*, vol. 126, no. 6, pp. 2356–2409, Dec. 2018.

[105] M. J. Tardy *et al.*, “Administrative office of the illinois courts pretrial operational review team,” Illinois Supreme Court Administrative Office of the Illinois Courts, Tech. Rep., Mar. 2014, p. 137.

[106] A. Ouss and M. Stevenson, *Does cash bail deter misconduct?* SSRN Scholarly Paper, Rochester, NY, Jan. 2022.

[107] A. Albright, “No money bail, no problems?” Tech. Rep., Nov. 2021, p. 61.

[108] D. Pager, R. Goldstein, H. Ho, and B. Western, “Criminalizing poverty: The consequences of court fees in a randomized experiment,” Tech. Rep., Mar. 2021, p. 41.

[109] L. Green, “Home is no castle for some cook county defendants; it’s jail,” *Injustice Watch*, Nov. 2016.

[110] C. Dizikes and T. Lightly, “Electronic monitoring spikes in cook county,” *Chicago Tribune*, Feb. 2015.

[111] P. Rogers, *More than 80 guns found in homes of individuals on electronic monitoring this year, cook county sheriff’s office says*, Jun. 2022.

[112] C. C. S. Office, *Cook county sheriff’s office community corrections - electronic monitoring (em) program (gps) information sheet*, Aug. 2020.

[113] CGL and C. Appleseed, “Electronic monitoring review cook county, illinois final report,” Cook County, IL, Report, Sep. 2022.

[114] R. Rosenfeld and A. Grigg, Eds., *The Limits of Recidivism: Measuring Success After Prison*. Washington, DC: The National Academies of Sciences, Engineering, and Medicine, 2022.

[115] S. Mello, “More cops, less crime,” *Journal of Public Economics*, vol. 172, pp. 174–200, Apr. 2019.

[116] J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz, “The rate of return to the highscope perry preschool program,” *Journal of Public Economics*, vol. 94, no. 1, pp. 114–128, Feb. 2010.

[117] I. Kohler-Hausmann, “Misdemeanor justice: Control without conviction,” *American Journal of Sociology*, vol. 119, no. 2, pp. 351–393, 2013.

[118] S. G. Mayson and M. Stevenson, “Misdemeanors by the numbers,” *Boston College Law Review*, vol. 61, no. 3, pp. 971–1044, Mar. 2020.

[119] V. Institute, *What jails cost: Cities chicago, il*, https://www.vera.org/publications/what-jails-cost-cities/chicago-il, 2022.

[120] S. J. Wilson and J. Lemoine, “Methods of calculating the marginal cost of incarceration: A scoping review,” *Criminal Justice Policy Review*, vol. 33, no. 6, pp. 639–663, Jul. 2022.

[121] B. S. Silveira, “Bargaining with asymmetric information: An empirical study of plea negotiations,” *Econometrica*, vol. 85, no. 2, pp. 419–452, 2017.

[122] P. Bayer, R. Hjalmarsson, and D. Pozen, “Building criminal capital behind bars: Peer effects in juvenile corrections,” *The Quarterly Journal of Economics*, vol. 124, no. 1, pp. 105–147, 2009.

[123] M. Stevenson, "Breaking bad: Mechanisms of social influence and the path to criminality in juvenile jails," *The Review of Economics and Statistics*, vol. 99, no. 5, pp. 824–838, Dec. 2017.

[124] J. L. Montiel Olea and C. Pflueger, "A robust test for weak instruments," *Journal of Business & Economic Statistics*, vol. 31, no. 3, pp. 358–369, Jul. 2013.

[125] G. W. Imbens and J. D. Angrist, "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 431–442, Jun. 1995.

[126] A. Fishbane, A. Ouss, and A. K. Shah, "Behavioral nudges reduce failure to appear for court," *Science*, vol. 370, no. 6517, Nov. 2020.

[127] A. Abid, *Taskforce, including preckwinkle, evans, and alvarez, visits washington d.c. to review pretrial system; and a warning about the expanded use of electronic monitoring*, Jul. 2014.

[128] A. D. Roy, "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, vol. 3, no. 2, pp. 135–146, 1951.

[129] G. W. Imbens and J. D. Angrist, "Identification and estimation of local average treatment effects," *Econometrica*, vol. 62, no. 2, pp. 467–475, 1994.

[130] E. Vytlacil, "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, vol. 70, no. 1, pp. 331–341, 2002.

[131] E. Vytlacil, "Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results," *The Review of Economics and Statistics*, vol. 88, no. 3, pp. 578–581, Aug. 2006.

[132] P. W. Holland, "Statistics and causal inference," *Journal of the American Statistical Association*, p. 27, 1986.

[133] P. Carneiro, J. Heckman, and E. Vytlacil, "Estimating marginal returns to education," *American Economic Review*, vol. 101, no. 6, pp. 2754–2781, Oct. 2011.

[134] J. Heckman and E. Vytlacil, "Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments," in *Handbook of Econometrics*, J. Heckman and E. E. Leamer, Eds., vol. 6, Elsevier, Jan. 2007, pp. 4875–5143.

[135] J. Grogger, "Certainty vs. severity of punishment," *Economic Inquiry*, vol. 29, no. 2, pp. 297–309, 1991.

[136] A. Bar-Ilan and B. Sacerdote, “The response to fines and probability of detection in a series of experiments,” National Bureau of Economic Research, Working Paper 8638, Dec. 2001.

[137] C. Mai and R. Subramanian, “The price of prisons: Examining state spending trends, 2010-2015,” Vera Institute for Justice, Tech. Rep., May 2017.

[138] J. Johnson, *June 16 2022 10:54 am i was called with a false violation in the middle of an interview*, Jun. 2022.

[139] D. McFadden, “Conditional logit analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, New York: Academic Press, 1974, pp. 105–142, ISBN: 978-0-12-776150-3.

[140] DOJ, “Advancing diversity in law enforcement report (october 2016),” DOJ Civil Rights Division and Equal Employment Opportunity Commission, The United States Department of Justice, Report, Oct. 2016.

[141] M. Keller, “Diversity on the force: Where police don’t mirror communities,” Governing, Tech. Rep., Sep. 2015, p. 14.

[142] M. Hoekstra and C. Sloan, “Does race matter for police use of force? evidence from 911 calls,” National Bureau of Economic Research, Working Paper 26774, Feb. 2020.

[143] F. Goncalves and S. Mello, “A few bad apples? racial bias in policing,” *American Economic Review*, vol. 111, no. 5, pp. 1406–1441, May 2021.

[144] B. Ba, D. Knox, J. Mummolo, and R. Rivera, “The role of officer race and gender in police-civilian interactions in chicago,” *Science*, vol. 371, no. 6530, pp. 696–702, Feb. 2021.

[145] E. K. Weisburst, ““whose help is on the way?” the importance of individual police officers in law enforcement outcomes,” Working Paper, Jul. 2020, p. 66.

[146] I. Ater, Y. Givati, and O. Rigbi, “Organizational structure, police activity and crime,” *Journal of Public Economics*, vol. 115, pp. 62–71, Jul. 2014.

[147] T. F. Pettigrew and L. R. Tropp, “A meta-analytic test of intergroup contact theory,” *Journal of Personality and Social Psychology*, vol. 90, no. 5, pp. 751–783, 2006.

[148] T. F. Pettigrew, “Secondary transfer effect of contact: Do intergroup contact effects spread to noncontacted outgroups?” *Social Psychology*, vol. 40, no. 2, p. 55, Jun. 2009.

[149] N. Tausch *et al.*, “Secondary transfer effects of intergroup contact: Alternative accounts and underlying processes.,” *Journal of Personality and Social Psychology*, vol. 99, no. 2, p. 282, Jul. 2010.

[150] G. A. Akerlof and R. E. Kranton, "Economics and identity," *The Quarterly Journal of Economics*, vol. 115, no. 3, pp. 715–753, Aug. 2000.

[151] R. G. Fryer and P. Torelli, "An empirical analysis of 'acting white'," *Journal of Public Economics*, vol. 94, no. 5, pp. 380–396, Jun. 2010.

[152] S. Anwar, P. Bayer, and R. Hjalmarsson, "Politics in the courtroom: Political ideology and jury decision making," *Journal of the European Economic Association*, vol. 17, no. 3, pp. 834–875, Jun. 2019.

[153] B. H. H. Golsteyn, A. Non, and U. Zölitz, "The impact of peer personality on academic achievement," *Journal of Political Economy*, vol. 129, no. 4, pp. 1052–1099, Apr. 2021.

[154] J. McCrary, "The effect of court-ordered hiring quotas on the composition and quality of police," *The American Economic Review*, vol. 97, no. 1, pp. 318–353, 2007.

[155] A. R. Miller and C. Segal, "Do female officers improve law enforcement quality? effects on crime reporting and domestic violence," *The Review of Economic Studies*, vol. 86, no. 5, pp. 2220–2247, Sep. 2018.

[156] J. J. Donohue and S. D. Levitt, "The impact of race on policing and arrests," *The Journal of Law & Economics*, vol. 44, no. 2, pp. 367–394, 2001.

[157] M. Garner, A. Harvey, and H. Johnson, "Estimating effects of affirmative action in policing: A replication and extension," *International Review of Law and Economics*, p. 105 881, Nov. 2019.

[158] A. Harvey and T. Mattia, "Reducing racial disparities in crime victimization," *Working Paper*, p. 45, Dec. 2019.

[159] R. Cox, J. P. Cunningham, and A. Ortega, "The impact of affirmative action litigation on police killings of civilians," Working Paper, 2021, p. 84.

[160] J. West, "Racial bias in police investigations," Working Paper, Oct. 2018, p. 37.

[161] E. Owens, D. Weisburd, K. L. Amendola, and G. P. Alpert, "Can you build a better cop?" *Criminology & Public Policy*, vol. 17, no. 1, pp. 41–87, 2018.

[162] B. Sacerdote, "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*, vol. 3, Elsevier, 2011, pp. 249–277, ISBN: 978-0-444-53429-3.

[163] C. Hoxby, "Peer effects in the classroom: Learning from gender and race variation," National Bureau of Economic Research, Working Paper 7867, Aug. 2000.

[164] B. Sacerdote, "Peer effects with random assignment: Results for dartmouth roommates," *The Quarterly Journal of Economics*, vol. 116, no. 2, pp. 681–704, 2001.

[165] J. D. Angrist, "The perils of peer effects," *Labour Economics*, Special Section Articles on "What Determined the Dynamics of Labour Economics Research in the Past 25 Years? Edited by Joop Hartog and and European Association of Labour Economists 25th Annual Conference, Turin, Italy, 19-21 September 2013 Edited by Michele Pellizzari, vol. 30, pp. 98–108, Oct. 2014.

[166] J. E. Holz, R. G. Rivera, and B. Ba, "Network effects in police use of force," Working Paper, Nov. 2019, p. 61.

[167] P. Ager, L. Bursztyn, L. Leucht, and H.-J. Voth, "Killer incentives: Rivalry, performance and risk-taking among german fighter pilots, 1939–45," *The Review of Economic Studies*, Dec. 2021.

[168] E. D. Gould, V. Lavy, and M. D. Paserman, "Does immigration affect the long-term educational outcomes of natives? quasi-experimental evidence," *The Economic Journal*, vol. 119, no. 540, pp. 1243–1269, 2009.

[169] V. Lavy and A. Schlosser, "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, vol. 3, no. 2, pp. 1–33, 2011.

[170] S. E. Black, P. J. Devereux, and K. G. Salvanes, "Under pressure? the effect of peers on outcomes of young adults," *Journal of Labor Economics*, vol. 31, no. 1, pp. 119–153, Jan. 2013.

[171] S. E. Carrell, M. Hoekstra, and E. Kuka, "The long-run effects of disruptive peers," *American Economic Review*, vol. 108, no. 11, pp. 3377–3415, Nov. 2018.

[172] A. A. Brenœ and U. Zölitz, "Exposure to more female peers widens the gender gap in stem participation," *Journal of Labor Economics*, vol. 38, no. 4, pp. 1009–1054, Oct. 2020.

[173] J. Boisjoly, G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles, "Empathy or antipathy? the impact of diversity," *The American Economic Review*, vol. 96, no. 5, pp. 1890–1905, 2006.

[174] S. E. Carrell, M. Hoekstra, and J. E. West, "The impact of college diversity on behavior toward minorities," *American Economic Journal: Economic Policy*, vol. 11, no. 4, pp. 159–182, Nov. 2019.

[175] S. Anwar, P. Bayer, and R. Hjalmarsson, "The impact of jury race in criminal trials," *The Quarterly Journal of Economics*, vol. 127, no. 2, pp. 1017–1055, 2012.

[176] D. Schindler and M. Westcott, “Shocking racial attitudes: Black g.i.s in europe,” *The Review of Economic Studies*, vol. 88, no. 1, pp. 489–520, Jan. 2021.

[177] T. F. Pettigrew, “Intergroup contact theory,” *Annual Review of Psychology*, vol. 49, no. 1, p. 21, 1998.

[178] C. V. Laar, S. Levin, S. Sinclair, and J. Sidanius, “The effect of university roommate contact on ethnic attitudes and behavior,” *Journal of Experimental Social Psychology*, vol. 41, no. 4, pp. 329–345, Jul. 2005.

[179] S. R. Sommers, “On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations,” *Journal of Personality and Social Psychology*, vol. 90, no. 4, pp. 597–612, 2006.

[180] S. Baker, A. Mayer, and S. L. Puller, “Do more diverse environments increase the diversity of subsequent interaction? evidence from random dorm assignment,” *Economics Letters*, vol. 110, no. 2, pp. 110–112, Feb. 2011.

[181] J. Burns, L. Corno, and E. L. Ferrara, “Interaction, prejudice and performance. evidence from south africa,” *Working Paper*, p. 52, Feb. 2015.

[182] L. P. Merlino, M. F. Steinhardt, and L. Wren-Lewis, “More than just friends? school peers and adult interracial relationships,” *Journal of Labor Economics*, vol. 37, no. 3, pp. 663–713, Jan. 2019.

[183] S. B. Billings, E. Chyn, and K. Haggag, “The long-run effects of school racial diversity on political identity,” *American Economic Review: Insights*, vol. 3, no. 3, pp. 267–284, Sep. 2021.

[184] D. Austen-Smith and R. G. Fryer Jr., “An economic analysis of “acting white”,” *The Quarterly Journal of Economics*, vol. 120, no. 2, pp. 551–583, May 2005.

[185] G. A. Akerlof and R. E. Kranton, *Identity Economics : How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton: Princeton University Press, 2010, ISBN: 978-0-691-15255-4.

[186] D. J. Benjamin, J. J. Choi, and A. J. Strickland, “Social identity and preferences,” *The American Economic Review*, vol. 100, no. 4, pp. 1913–1928, 2010.

[187] R. Holden, M. Keane, and M. Lilley, “Peer effects on the united states supreme court,” Working Paper, Jan. 2021, p. 47.

[188] CPD, *Cpd 2017 faq*, 2017.

[189] CPD, *Education and training division (etd) | chicago police department*, 2020.

[190] CPD, *Field training and evaluation program*, http://directives.chicagopolice.org/directives/data/a7a57be2-1294231a-bf312-942c-e1f46fde5fd8c4e8.html?hl=true, Jun. 2018.

[191] P. Pritchard, "Do you have what it takes to join the chicago police department?" *Chicago magazine*, Aug. 2013.

[192] feredeathpsn, *Chicago police lottery number*, Reddit Post, Jun. 2017.

[193] CPD, *Personnel transfer and assignment procedures –(fop)*, http://directives.chicagopolice.org/directives/c 12bcf25e-31612-bcf2-5ebc1c9f5d96947f.html?hl=true, Dec. 2011.

[194] A. Mas, "Pay, reference points, and police performance," *The Quarterly Journal of Economics*, vol. 121, no. 3, pp. 783–821, 2006.

[195] L. Shi, "The limit of oversight in policing: Evidence from the 2001 cincinnati riot," *Journal of Public Economics*, vol. 93, no. 1-2, pp. 99–113, Feb. 2009.

[196] D. Coviello and N. Persico, "An economic analysis of black-white disparities in the new york police department's stop-and-frisk program," *The Journal of Legal Studies*, vol. 44, no. 2, pp. 315–360, 2015.

[197] J. Blanes i Vidal and T. Kirchmaier, "The effect of police response time on crime clearance rates," *The Review of Economic Studies*, vol. 85, no. 2, pp. 855–891, Apr. 2018.

[198] T. Kirchmaier, S. Machin, M. Sandi, and R. Witt, "Joining forces? crewing size and the productivity of policing," Working Paper, Mar. 2021, p. 62.

[199] R. Chetty, J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan, "How does your kindergarten classroom affect your earnings? evidence from project star," *The Quarterly Journal of Economics*, vol. 126, no. 4, pp. 1593–1660, 2011.

[200] C. F. Manski, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, vol. 60, no. 3, pp. 531–542, 1993.

[201] D. Card and A. B. Krueger, "Does school quality matter? returns to education and the characteristics of public schools in the united states," *Journal of Political Economy*, vol. 100, no. 1, pp. 1–40, Feb. 1992.

[202] B. R. Close and P. L. Mason, "Searching for efficient enforcement: Officer characteristics and racially biased policing," *Review of Law and Economics*, vol. 3, no. 2, pp. 263–322, 2007.

[203] L. R. Bergé, "Efficient estimation of maximum likelihood models with multiple fixed-effects: The r package fenmlm," *CREA Discussion Papers*, no. 13, p. 39, 2018.

[204] D. Acemoglu and J. Angrist, “How large are human-capital externalities? evidence from compulsory schooling laws,” *NBER Macroeconomics Annual*, vol. 15, pp. 9–59, Jan. 2000.

[205] J. Feld and U. Zölitz, “Understanding peer effects: On the nature, estimation, and channels of peer effects,” *Journal of Labor Economics*, Jan. 2017.

[206] J. D. Angrist and K. Lang, “Does school integration generate peer effects? evidence from boston’s metco program,” *American Economic Review*, vol. 94, no. 5, pp. 1613–1634, Dec. 2004.

[207] S. A. Imberman, A. D. Kugler, and B. I. Sacerdote, “Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees,” *American Economic Review*, vol. 102, no. 5, pp. 2048–2082, May 2012.

[208] S. E. Carrell, B. I. Sacerdote, and J. E. West, “From natural variation to optimal policy? the importance of endogenous peer group formation,” *Econometrica*, vol. 81, no. 3, pp. 855–882, 2013.

[209] B. Caeyers and M. Fafchamps, “Exclusion bias in the estimation of peer effects,” National Bureau of Economic Research, Working Paper 22565, Aug. 2016.

[210] S. Athey and G. Imbens, “The econometrics of randomized experiments,” *arXiv:1607.00698 [econ, stat]*, Jul. 2016. arXiv: `1607.00698 [econ, stat]`.

[211] J. Roth and P. H. C. Sant’Anna, “Efficient estimation for staggered rollout designs,” Working Paper, Jun. 2021. arXiv: `2102.01291`.

[212] E. H. Chang *et al.*, “The mixed effects of online diversity training,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 16, pp. 7778–7783, Apr. 2019.

[213] F. Dobbin and A. Kalev, “Why diversity programs fail,” *Harvard Business Review*, vol. 94, no. 7, 2016.

[214] G. Cohen, “Public administration training in basic police academies: A 50-state comparative analysis,” *The American Review of Public Administration*, vol. 51, no. 5, pp. 345–359, Jul. 2021.

[215] M. D. Webb, “Reworking wild bootstrap based inference for clustered errors,” p. 23, Nov. 2014.

[216] J. M. MacDonald, “De-policing as a consequence of the so-called "ferguson effect",” *Criminology & Public Policy*, vol. 18, no. 1, pp. 47–49, 2019.

[217] E. Owens, “Economic approach to "de-policing",” *Criminology & Public Policy*, vol. 18, no. 1, pp. 77–80, 2019.

[218] H. Mac Donald, "The war on cops," *New York: Encounter Books*, 2016.

[219] C. Prendergast, "Selection and oversight in the public sector, with the los angeles police department as an example," National Bureau of Economic Research, Tech. Rep., 2001.

[220] L. Shi, "The limit of oversight in policing: Evidence from the 2001 cincinnati riot," *Journal of Public Economics*, vol. 93, no. 1, pp. 99–113, 2009.

[221] R. Cassell Paul G. Fowles, "What caused the 2016 chicago homicide spike: An empirical examination of the aclu effect and the role of stop and frisks in preventing gun violence symposium: Federal responses to police misconduct: Possibilities and limits," *University of Illinois Law Review*, vol. 2018, p. 1581, 2018.

[222] D. Premkumar, "Public Scrutiny and Police Effort: Evidence from Arrests and Crime After High-Profile Police Killings," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3715223, Sep. 2019.

[223] T. Devi and J. Fryer Roland G, "Policing the Police: The Impact of "Pattern-or-Practice" Investigations on Crime," National Bureau of Economic Research, Working Paper 27324, Jun. 2020, Series: Working Paper Series.

[224] C. Prendergast, "'Drive and Wave': The Response to LAPD Police Reforms After Rampart," Working Paper, 2021.

[225] P. Ganong, F. Greig, P. Noel, D. M. Sullivan, and J. Vavra, "Micro and macro disincentive effects of expanded unemployment benefits," *Working Paper*, 2021.

[226] C. F. Manski and J. V. Pepper, "Deterrence and the death penalty: Partial identification analysis using repeated cross sections," *Journal of quantitative criminology*, vol. 29, no. 1, pp. 123–141, 2013.

[227] C. F. Manski and J. V. Pepper, "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *The Review of Economics and Statistics*, vol. 100, no. 2, pp. 232–244, May 2018.

[228] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program," *Journal of the American statistical Association*, vol. 105, no. 490, pp. 493–505, 2010.

[229] D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager, "Synthetic difference in differences," *American economic review*, vol. Forthcoming, 2021.

[230] F. X. Lee and W. Suen, "Credibility of crime allegations," *American Economic Journal: Microeconomics*, 2020.

[231] S. D. Levitt, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, vol. 87, no. 3, pp. 270–290, 1997.

[232] W. N. Evans and E. G. Owens, "COPS and crime," *Journal of Public Economics*, vol. 91, no. 1-2, pp. 181–201, 2007.

[233] C. Fu and K. I. Wolpin, "Structural Estimation of a Becker-Ehrlich Equilibrium Model of Crime: Allocating Police Across Cities to Reduce Crime," *The Review of Economic Studies*, vol. 85, no. 4, pp. 2097–2138, Nov. 2017.

[234] A. Chalfin and J. McCrary, "Are us cities underpoliced? theory and evidence," *Review of Economics and Statistics*, vol. 100, no. 1, pp. 167–186, 2018.

[235] S. Mello, "More COPS, less crime," *Journal of Public Economics*, vol. 172, pp. 174–200, Apr. 2019.

[236] R. Di Tella and E. Schargrodsky, "Do police reduce crime? estimates using the allocation of police forces after a terrorist attack," *American Economic Review*, vol. 94, no. 1, pp. 115–133, 2004.

[237] M. Draca, S. Machin, and R. Witt, "Panic on the streets of london: Police, crime, and the july 2005 terror attacks," *American Economic Review*, vol. 101, no. 5, pp. 2157–81, 2011.

[238] J. Klick and A. Tabarrok, "Using terror alert levels to estimate the effect of police on crime," *The Journal of Law and Economics*, vol. 48, no. 1, pp. 267–279, 2005.

[239] J. Blanes i Vidal and T. Kirchmaier, "The effect of police response time on crime clearance rates," *The Review of Economic Studies*, vol. 85, no. 2, pp. 855–891, 2017.

[240] P. Heaton, P. Hunt, J. MacDonald, and J. Saunders, "The short- and long-run effects of private law enforcement: Evidence from university police," *The Journal of Law and Economics*, vol. 59, no. 4, pp. 889–912, 2016.

[241] S. Weisburd, "Police presence, rapid response rates, and crime prevention," *Working Paper*, 2018.

[242] J. B. i Vidal and G. Mastrobuoni, "Police patrols and crime," *Working Paper*, 2018.

[243] G. S. Becker and G. J. Stigler, "Law enforcement, malfeasance, and compensation of enforcers," *The Journal of Legal Studies*, pp. 1–18, 1974.

[244] J. P. Cunningham and R. Gillezeau, "Don't shoot! the impact of historical african american protest on police killings of civilians," *Journal of Quantitative Criminology*, 2019.

[245] S. Gaston, J. P. Cunningham, and R. Gillezeau, "A ferguson effect, the drug epidemic, both, or neither? explaining the 2015 and 2016 u.s. homicide rises by race and ethnicity," *Homicide Studies*, vol. 23, no. 3, pp. 285–313, 2019.

[246] D. S. Kirk and M. Matsuda, "Legal cynicism, collective efficacy, and the ecology of arrest," *Criminology*, vol. 49, no. 2, pp. 443–472, 2011.

[247] M. Desmond, A. V. Papachristos, and D. S. Kirk, "Police violence and citizen crime reporting in the black community," *American Sociological Review*, vol. 81, no. 5, pp. 857–876, 2016.

[248] M. D. Reisig, S. E. Wolfe, and K. Holtfreter, "Legal cynicism, legitimacy, and criminal offending: The nonconfounding effect of low self-control," *Criminal Justice and Behavior*, vol. 38, no. 12, pp. 1265–1279, 2011.

[249] D. Ang, "The effects of police violence on inner-city students," *Working paper*, 2018.

[250] J. Mummolo, "Modern police tactics, police-citizen interactions and the prospects for reform," *The Journal of Politics*, vol. 80, no. 1, pp. 1–15, 2018.

[251] C. Cheng and W. Long, "Improving police services: Evidence from the French Quarter Task Force," *Journal of Public Economics*, vol. 164, no. C, pp. 1–18, 2018.

[252] N. Mastrorocco and A. Ornaghi, "Who Watches the Watchmen? Local News and Police Behavior in the United States," Trinity College Dublin, Department of Economics, Tech. Rep. tep0720, Nov. 2020, Publication Title: Trinity Economics Papers.

[253] C. F. Manski and D. S. Nagin, "Assessing benefits, costs, and disparate racial impacts of confrontational proactive policing," *Proceedings of the National Academy of Sciences*, 2017.

[254] J.-P. Benoit and J. Dubra, "Why do good cops defend bad cops?" *International Economic Review*, vol. 45, no. 3, pp. 787–809, 2004.

[255] B. A. Ba, "Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in chicago," *Working paper*, 2018.

[256] D. Henstock and B. Ariel, "Testing the effects of police body-worn cameras on use of force during arrests: A randomised controlled trial in a large british police force," *European Journal of Criminology*, vol. 14, no. 6, pp. 720–750, 2017.

[257] B. Ariel, W. A. Farrar, and A. Sutherland, "The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial," *Journal of quantitative criminology*, vol. 31, no. 3, pp. 509–535, 2015.

[258] B. Ariel *et al.*, “Report: Increases in police use of force in the presence of body-worn cameras are driven by officer discretion: A protocol-based subgroup analysis of ten randomized experiments,” *Journal of Experimental Criminology*, vol. 12, no. 3, pp. 453–463, 2016.

[259] B. Ariel *et al.*, “Wearing body cameras increases assaults against officers and does not reduce police use of force: Results from a global multi-site experiment,” *European Journal of Criminology*, vol. 13, no. 6, pp. 744–755, 2016.

[260] B. Ariel, “Increasing cooperation with the police using body worn cameras,” *Police Quarterly*, vol. 19, no. 3, pp. 326–362, 2016.

[261] B. Ariel *et al.*, “Contagious accountability a global multisite randomized controlled trial on the effect of police body-worn cameras on citizens complaints against the police,” *Criminal justice and behavior*, vol. 44, no. 2, pp. 293–316, 2017.

[262] D. Henstock and B. Ariel, “Testing the effects of police body-worn cameras on use of force during arrests: A randomised controlled trial in a large british police force,” *European Journal of Criminology*, vol. 14, no. 6, pp. 720–750, 2017.

[263] D. Yokum, A. Ravishankar, and A. Coppock, “A randomized control trial evaluating the effects of police body-worn cameras,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 21, pp. 10 329–10 332, 2019.

[264] T. Kim, “The impact of body worn cameras on police use of force and productivity: Evidence from new jersey,” *Working Paper*, 2020.

[265] S. N. Durlauf, “Assessing racial profiling,” *The Economic Journal*, vol. 116, no. 515, F402–F426, 2006.

[266] C. F. Manski, “Optimal search profiling with linear deterrence,” *The American Economic Review*, vol. 95, no. 2, pp. 122–126, 2005.

[267] C. F. Manski, “Search profiling with partial knowledge of deterrence,” *The Economic Journal*, vol. 116, no. 515, F385–F401, 2006.

[268] J. Knowles, N. Persico, and P. Todd, “Racial bias in motor vehicle searches: Theory and evidence,” *Journal of Political Economy*, vol. 109, no. 1, pp. 203–229, 2001.

[269] S. Anwar and H. Fang, “An alternative test of racial prejudice in motor vehicle searches: Theory and evidence,” *American Economic Review*, vol. 96, no. 1, pp. 127–151, 2006.

[270] K. Antonovics and B. G. Knight, “A new look at racial profiling: Evidence from the boston police department,” *The Review of Economics and Statistics*, vol. 91, no. 1, pp. 163–177, 2009.

[271] D. Dharmapala and T. J. Miceli, "Search, Seizure and (False?) Arrest: An Analysis of Fourth Amendment Remedies when Police can Plant Evidence," *Working Paper*, 2012.

[272] D. Ang, P. Bencsik, J. Bruhn, and E. Derenoncourt, "Police Violence Reduces Civilian Cooperation and Engagement with Law Enforcement," Working Paper, 2021.

[273] J. Kaplan. "Uniform crime reporting program data: Offenses known and clearances by arrest, 1960-2016." (2018).

[274] A. H. Wu, "Gender Bias among Professionals: An Identity-Based Interpretation," *The Review of Economics and Statistics*, vol. 102, no. 5, pp. 867–880, Dec. 2020.

[275] P. Heaton, "Understanding the effects of antiprofiling policies," *The Journal of Law and Economics*, vol. 53, no. 1, pp. 29–64, 2010.

[276] D. Kidwell, *Appeals court declares police internal affairs files public records - Sunshine Online*, Jul. 2009.

[277] D. S. Nagin and G. Pogarsky, "Integrating Celerity, Impulsivity, and Extralegal Sanction Threats into a Model of General Deterrence: Theory and Evidence*," *Criminology*, vol. 39, no. 4, pp. 865–892, 2001, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-9125.2001.tb00943.x.

[278] M. Davey and M. Smith, "Video of chicago police shooting a teenager is ordered released," *The New York Times*, 2015.

[279] J. J. Donohue, "Comey, trump, and the puzzling pattern of crime in 2015 and beyond," *Columbia Law Review*, vol. 117, no. 5, pp. 1297–1354, 2017.

[280] A. Abadie and J. Gardeazabal, "The economic costs of conflict: A case study of the basque country," *American economic review*, vol. 93, no. 1, pp. 113–132, 2003.

[281] A. Abadie, A. Diamond, and J. Hainmueller, "Comparative politics and the synthetic control method," *American Journal of Political Science*, vol. 59, no. 2, pp. 495–510, 2015.

[282] B. Ba, P. Bayer, N. Rim, R. Rivera, and M. Sidibe, "Police officer assignment mechanisms and neighborhood crime," *Working Paper*, 2021.

[283] C. Lum and D. S. Nagin, "Reinventing american policing," *Crime and justice*, vol. 46, no. 1, pp. 339–393, 2017.

[284] Department of Justice, "Investigation of the Chicago Police Department," Civil Rights Office, Tech. Rep., 2017.

[285] P. Leifeld, "Texreg: Conversion of statistical model output in r to latex and html tables," *Journal of Statistical Software*, vol. 55, pp. 1–24, Nov. 2013.

[286] T. C. Federation, "The impact of cook county bond court on the jail population: A call for increased public data and analysis," The Civic Federation, Chicago, Tech. Rep., Nov. 2017.

[287] T. C. Federation, *Cook county seeks consulting services to review electronic monitoring practices | the civic federation*, May 2020.

[288] J. D. Angrist, G. W. Imbens, and A. B. Krueger, "Jackknife instrumental variables estimation," *Journal of Applied Econometrics*, vol. 14, no. 1, pp. 57–67, 1999.

[289] P. Goldsmith-Pinkham, P. Hull, and M. Kolesár, "Contamination bias in linear regressions," Working Paper, Jun. 2022, p. 46.

[290] J. Afeef, L. Bostwick, S. Kim, and J. Reichert, "Policies and procedures of the criminal justice system | office of justice programs," Illinois Criminal Justice Information Authority, Tech. Rep. 247298, Aug. 2012, p. 45.

[291] M. Mogstad, A. Santos, and A. Torgovitsky, "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, vol. 86, no. 5, pp. 1589–1619, 2018.

[292] G. W. Imbens and D. B. Rubin, "Estimating outcome distributions for compliers in instrumental variables models," *The Review of Economic Studies*, vol. 64, no. 4, pp. 555–574, Oct. 1997.

[293] A. Abadie, "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, vol. 113, no. 2, pp. 231–263, Apr. 2003.

[294] M. Bhuller, G. B. Dahl, K. V. L, and M. Mogstad, "Incarceration spillovers in criminal and family networks," National Bureau of Economic Research, Working Paper 24878, Aug. 2018, p. 29.

[295] J. M. Ferguson and D. Witzburg, "Evaluation of the demographic impacts of the chicago police department's hiring process," Office of the Inspector General, City of Chicago, Report, 2021, p. 70.

[296] J. Kass and R. Blau, "Police hiring lottery latest daley headache," *Chicago Tribune*, p. 3, Aug. 1991.

[297] Chicago_mwk, *Chicago police academy 2010*, Jan. 2010.

[298] neverlose357, *2011 chicago police academy*, Dec. 2010.

[299] Aendos, *Chicago police 2016*, Aug. 2015.

[300] J. Guryan, K. Kroft, and M. J. Notowidigdo, “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments,” *American Economic Journal: Applied Economics*, vol. 1, no. 4, pp. 34–68, 2009.

[301] R. Chetty, J. N. Friedman, and J. E. Rockoff, “Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates,” *American Economic Review*, vol. 104, no. 9, pp. 2593–2632, Sep. 2014.

[302] C. N. Morris, “Parametric empirical bayes inference: Theory and applications,” *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 47–55, 1983.

[303] J. Hahn and Z. Liao, “Bootstrap standard error estimates and inference,” *Econometrica*, vol. 89, no. 4, pp. 1963–1977, 2021.

[304] R. A. Fisher, “Theory of statistical estimation,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 22, no. 5, pp. 700–725, Jul. 1925.

[305] S. Athey, D. Eckles, and G. W. Imbens, “Exact p-values for network interference,” *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 230–240, Jan. 2018.

[306] V. Michelman, J. Price, and S. D. Zimmerman, “Old boys’ clubs and upward mobility among the educational elite,” *The Quarterly Journal of Economics*, Dec. 2021.

[307] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, no. 2, Feb. 2004.

[308] G. Csardi and T. Nepusz, “The igraph software package for complex network research,” *InterJournal*, vol. Complex Systems, p. 1695, Nov. 2005.

[309] M. D. Maltz and H. E. Weiss, “Creating a ucr utility: Final report to the national institute of justice,” *NIJ Research Report*, vol. 15341, pp. 1–21, 2006.

[310] J. S. Neyman, “On the application of probability theory to agricultural experiments. essay on principles. section 9.(translated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480),” *Annals of Agricultural Sciences*, vol. 10, pp. 1–51, 1923.

[311] D. B. Rubin, “Estimating causal effects of treatments in randomized and nonrandomized studies.,” *Journal of educational Psychology*, vol. 66, no. 5, p. 688, 1974.

[312] P. W. Holland, “Statistics and causal inference,” *Journal of the American statistical Association*, vol. 81, no. 396, pp. 945–960, 1986.

Appendix A: Chapter 1

A.1 Additional Analyses and Background

A.1.1 Background

The judge can also link I and D bonds with supervised release requirements, though the main role of the bond is to determine if they can leave the custody of the Sheriff (i.e., exit jail pretrial). In this sense, EM can also be coupled with D-bonds (D-EM), which required the defendant to stay in jail until they paid 10% of the bond amount and were then released onto EM, which accounted for 17% of D-bonds between 2008 and 2012 ([286]). However, it is unclear how many defendants were actually released onto EM from D-EMs during the period. Prior to 2012, little data on EM usage in Cook County was available ([110]). Figure A.3 displays these trends across the sample period.

In 2012, disputes began between the Court and the Sheriff (who runs the jail and most of the EM releases) over the overcrowding of the Cook County Jail and EM usage began. As a result, in November 2012, judges functionally stopped using D-EM bonds which further contributed to jail overcrowding ([286]), though they were occasionally used during the period of study. This sparked the introduction of the IEM bond discussed in the paper, though IEM bonds are referred to as "Electronic monitoring with D-bonds" in [113]. The IEM bond offered an attractive solution to judges: release defendants from jail and avoid overcrowding but have them monitored by the Sheriff, who bears responsibility for any failures. The initial appeal was increased following reforms in September 2013 which urged a reduction in defendants forced to stay in jail due to lack of money ([286]). A report by the Civic Federation, using different data, also indicates that in September 2013, IEM was about 25% of dispositions ([286]).

A significantly less common EM program was "Curfew EM" which requires defendants to be in their homes between specified hours, usually 7 pm to 7 am. These programs also co-exist with other monitored release programs by the Chief Judge's office, GPS home confinement, which is primarily used for domestic violence cases ([287]). See here for a discussion of differences between EM programs.

Recently, Cook County has adopted GPS monitoring systems instead, though these GPS ankle or wrist bracelets operate in a similar capacity, simply without a home unit, and tracks all of the subject's movements (Sheriff (2020)).

A.1.2 Data

The court data has large numbers of cases, first starting in 1984, but also contains sporadic records dating back to the 1930s. Linking is done using individual record numbers as well as personally identifiable characteristics, such as name, birth date, race, gender, and home address. As a single booking can result in multiple cases (generally 2 if it is a felony case), cases can be linked within individuals using central booking numbers common to both cases (RD numbers if CBs are missing). For linking jail data to court data, I connect defendant identities using individual record numbers, identifiable information (names), and case/detention information.

The CB and RD numbers associated with cases are used to connect court/jail profiles to Chicago-specific arrests and reported crimes. While I have additional information for Chicago arrestees, I do not require this filter, though 92.37% of the data are reported to have been arrested by the CPD — 87.32% can be linked to a specific arrest, while 73.47% can be linked to a specific crime.

Importantly, not all cases can be linked to a reasonable jail spell, which means that the individual follows a quick timeline of beginning a jail spell (i.e., defendant is reported to have entered jail), having a case opened against them, and proceeding to bond court. This lack of linkage is possible due to some individuals never entering the jail system due to immediately going to bond court and being released or due to a linking error — I test the sensitivity of my results to these unlinked cases in Section 2.5. Cases with I-bonds or EM-bonds are much more likely to be unlinked to a jail spell (61.32% and 25.08%, respectively) relative to all other bond types (which averages

13.81%), which supports the former case. Interestingly, there is no pattern in missing rates for increasing bond amounts. If this were largely due to immediate release from EM, we would expect higher bonds to imply fewer missing. For instrument construction, I include all cases; for treatment construction in the main specification, non-I-bond and non-EM cases which are unlinked to jail spells are dropped in the main sample, while unlinked I-bond and EM bond cases are kept. A defendant can also be classified as "detained" if they technically exit jail due to a transfer or are sent to alternative detention (e.g., prison).

For final filters, I exclude cases that went the branch 1 within 2 days of being opened. I also remove a small subset of individual-booking observations with irregular case patterns and those which do not have resolutions within the court system. I remove 1.94% if they contain more than 3 unique cases, if there are multiple cases and the difference between the minimum and maximum case initialization date is more than 120 days, if the defendant had more than 60 past cases, and if there were more than 6 individual-bookings corresponding to that defendant within the 2 year period. I drop cases within this time period that are transferred outside of the regular system (1.61%), have short case histories without resolution (3.19%), or end with a warrant being issued (0.24%).

Some cases do not have final disposition dates but end with the case being dropped and contain a guilty, not guilty, stricken, or dropped disposition code (4.91%), I use the last event date as the final disposition date. Lastly, I remove a small number of cases that had murder or felony sex charges or resulted in bond denial, and I drop cases without a categorizable treatment (which includes missing jail spells for defendants without I or EM bonds). In Section 2.5, I test the sensitivity of the results to alternate classifications of the dropped cases due to missing jail information.

In July 2015, the court introduced a public safety assessment system that guided judges on release decisions using a scoring system ([287]). However, it is not clear in the data if this influenced judge behavior in any way. After 2015 following defendants into future cases becomes more difficult due to the data ending in 2019, and there were significant changes in Chicago in 2016. As a result, I limit the data to July 2013 through 2015.

A.1.3 Tests for Interpreting 2SLS Results as LATEs

As shown in [98], under certain conditions, we can recover a LATE for one treatment level versus an adjacent one (EM versus release and EM versus detention) with 2SLS. In particular, there are two sets of assumptions relevant for this paper.

Linearity of Predicted Treatments Assuming a single index crossing multiple thresholds model, as presented in Section 1.4, is correct, then we require that predicted treatment probabilities are linear functions of each other — for example, $\mathbb{E}[Z^1|Z^3]$ is linear in Z^3. Figure A.6 displays the results for the informal test of this linearity assumption. It shows that, by comparing a linear fit to the local polynomial fit, there are some mild deviations from linearity in the full sample, but in the felony sample the fit is near-linear for the entire distribution. However, the assumption applies globally to Z^1 and Z^3 and so even the minor deviations from the linear fit in the felony sample at the tails constitute a small violation.

Average Monotonicity and No Cross Effects Without additional assumptions on selection into treatment, we require two conditions for 2SLS to recover a LATE: average monotonicity (e.g., there is a positive correlation between $Release_{ic}$ and Z_1) across sub-samples and no cross effects (e.g., there is a no correlation between $Release_{ic}$ and Z_3, conditional on Z_1). We can test for violations of average monotonicity across sub-samples and of no cross effects. Table A.6 displays the results of these tests. Across sub-samples, the endogenous treatments are strongly and positively correlated with their respective instrument, and across sub-samples the relationship of a treatment with the unrelated instrument, conditional on the correct instrument, is either statistically or economically insignificant (e.g., the coefficient for Z_a is at a minimum 6 times larger in magnitude than that of Z_b with treatment a as the dependent variable). Nevertheless, these do constitute violations of the no cross effect assumptions.

Overall, these results do constitute violations of the assumptions, but in both cases they are relatively minor. However, as of this writing, the tests in [98] are not formalized, nor do we know to what degree minor violations bias results. With this, I cautiously proceed with the interpretation of the 2SLS results as a LATE.

A.1.4 Construction of Treatment Parameters

Because I assume MTRs are additively separable in u and x, the construction of the observable and unobservable components are done separately, using a uniform grid of 99 points ($u \in \{0.01, 0.02, ..., 0.99\}$), following the construction of the average treatment effect on the treated (for ATR, ATD) and ATE from [99].

$$ATE_{2,1} = \frac{1}{N}\sum_{i=1}^{N} X_i(\beta_2 - \beta_1) + \frac{1}{n_{p_1}} \sum_{u=100\times\underline{p_1}}^{100\times\overline{p_1}} E[\omega_2 - \omega_1 | u = u]$$

$$ATE_{2,3} = \frac{1}{N}\sum_{i=1}^{N} X_i(\beta_2 - \beta_3) + \frac{1}{n_{p_2}} \sum_{u=100\times\underline{p_2}}^{100\times\overline{p_2}} E[\omega_2 - \omega_3 | u = u]$$

$$ATR = \frac{1}{N}\sum_{i=1}^{N} \frac{p_{1,i}}{\hat{p_1}} X_i(\beta_2 - \beta_1) + \frac{1}{n_{p_1}} \sum_{u=100\times\underline{p_1}}^{100\times\overline{p_1}} \frac{\Pr(p_1 > 1 - \frac{u}{100})}{\hat{p_1}} E[\omega_2 - \omega_1 | u = u]$$

$$ATD = \frac{1}{N}\sum_{i=1}^{N} \frac{p_{2,i}}{\hat{p_2}} X_i(\beta_2 - \beta_3) + \frac{1}{n_{p_2}} \sum_{u=100\times\underline{p_2}}^{\overline{p_2}} \frac{\Pr(p_2 > \frac{u}{100})}{\hat{p_2}} E[\omega_2 - \omega_3 | u = u]$$

where N is the number of observations (cases), $p_1 = 1 - \pi_1$, $p_2 = \pi_2$, $\overline{x}$ and $\underline{x}$ refer to the upper and lower limits of common support, n_x is the number of points between the upper and lower limits, and $\hat{x}$ is the average over the range of common support.

A.1.5 Comparison with Prior Methods

Figure A.24 displays MTEs within common support ("Main"), polynomial MTEs with full support using sieve-style estimation ("Poly"), and Normal method MTEs for a subset of outcomes. Noticeably, when the MTEs are linear with Main and Poly (e.g., in the case of FTAs or sentence to incarceration for EM versus release), they match the Normal MTEs closely for most values of u, for which the Normal MTEs are linear by construction. At more extreme values of u, however, the Normal MTEs' pre-determined shape leads to more extreme values.

The patterns diverge more significantly when the MTEs for Main and Poly are non-linear (in most cases), leading the Normal MTEs to extrapolate too far different results or mask underlying heterogeneity. For example, using EM versus detention: for the likelihood of having a new case pretrial, Main and Poly produce a parabolic shape, while the Normal method produces a flat line; for the likelihood of being sentenced to incarceration and total case cost over 4 years, Main and Poly have a parabolic shape resulting in near zero effects on high u defendants while the Normal estimates continue moving negative implying large effects for high u defendants.

A.1.6 Additional Robustness Checks

Alternative Specifications

Results are displayed in Figure A.25.

Alternative π Computation The main results use π_1 and π_2 based on equation (1.1) using a probit model. I test the robustness of this using two modifications. First, because I do not exclude defendants' observations from the data I use to construct their judge-specific predicted propensities (πs), we may be concerned that defendant unobservables could be biasing the instruments. Though this is unlikely due to the fact that each judge sees thousands of cases and defendants, I test the robustness of the results with respect to this concern by re-calculating $\pi_{1,i}$ and $\pi_{2,i}$ using only observations excluding defendant i, then predicting out of sample, similar to a jackknife instrument ([288]) used in [30].

Second, [289] show that judge-instrument first stages can suffer from contamination bias. One recommended solution is to interact judges with fixed effects for the level of randomization (e.g., court rooms). I this paper, judges are randomly assigned conditional on time fixed effects. While not directly applicable, as I use a probit first stage, not OLS, and interact judges with defendant observables, I add judge-time interactions to my first stage (equation (1.1)) to test the robustness of my results. The results for both of these tests labeled 'Out-Sample' and 'Interact FEs', respectively, and the results are largely the same.

Alternate Specifications The main results use 3rd degree polynomials for Φ_s. To test the sensitivity of the results with respect to this specification, I re-estimate the MTRs and construct MTEs under 3 alternative specifications for Φ_s: 3rd degree B-splines with 4 degrees of freedom, 4th degree B-splines with 6 degrees of freedom, and a 5th degree polynomial. For each of these alternative specifications, the results are generally similar to the main results.

Alternative Samples

Results are displayed in Figure A.26.

Bonds as Treatments It is useful to know how sensitive the results are to mismeasuring the treatment. So, rather than determining treatment based on a day-in-jail cutoff which uses both court and jail data, I recode the treatments to be based on bonds: release is now only I-Bonds and low D-bonds ($<$ \$20,000), EM is all EM bonds, and detention is all D-bonds $\geq$ \$20,000. The results are generally consistent with the main results. For new case pretrial, the pattern is more exaggerated with larger increases for low-resistance and a modest decrease for mid-level defendants for EM versus release, and for detention versus EM the curve is shifted downward, displaying a larger incapacitation effect of detention. for Sentences to incarceration for detention versus EM the pattern is flat and sloping upwards (but still negative).

CPD Rearrests Instead of using court records to determine new pretrial cases, I can use CPD arrest records instead. For this, I limit the sample to the CPD arrestees (making this sample similar to that of housing-secure and CPD-only samples). The MTEs for CPD rearrests are shown in the same figure as new cases pretrial. The results for EM versus release are highly similar to the main results, though the MTE for detention versus EM is shifted upwards for a net negative effect.

Chicago Arrests Only The main results may be altered because I only have data on arrests and crimes within Chicago which allows for improved matching earlier on in the data cleaning, which is only a part of Cook County. To see if this is significantly influencing my results, I subset my analysis to cases that were initiated by arrests by the Chicago Police Department (92.74% of the felony cases). The results are similar to the main results.

Excluding Non-Home Addresses There is also the potential that the treatment contains selection bias because individuals without stable living conditions cannot be released on EM or from jail if they have no residence to go to — alternatively, individuals without stable living environments may also be forced to violate their EM conditions. In order to determine if this is influencing my results, I subset the data to individuals for whom I have address information for their case. This keeps cases with addresses within the bounds of Chicago and excludes cases if the address is associated with a homeless shelter, recovery facility, or halfway house. This leaves 77.15% of the full sample observations. Overall, for almost all outcomes, the results tract very closely to the CPD-only results and are generally consistent with the patterns and conclusions of the main sample.

Including Misdemeanor Cases Rather including only felony cases, I include misdemeanor cases and D-EM bonds in estimating the MTEs. The misdemeanor sample by itself lacks significant support between treatments, so these results are suggestive at best. Overall, the results are similar to the main sample result, with a few exceptions, particularly for EM versus release: the effect of EM relative to release on FTAs is larger, the new case pretrial pattern is flipped (large negative effects for low resistance defendants which moves positive), and the effect on new case cost is downward sloping.

A.2 Additional Figures and Tables

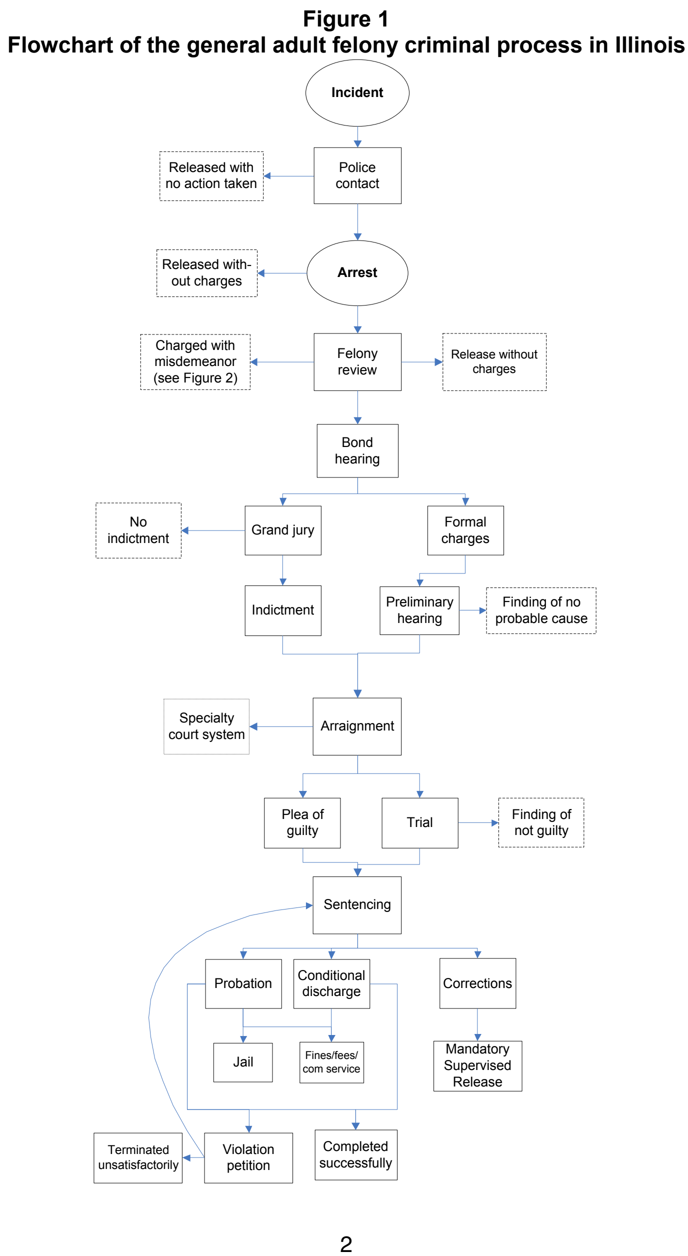


Figure A.1: Felony Case Flow Chart

Note: Figure displays sequence of events for felony cases within Cook County — though document is meant for the entire Illinois criminal justice system more broadly. Source: [290], page 2.

	Release (1)	EM (2)	Detain (3)
Median Days in Jail	1	-	92
Median Days on EM	0	42	-
Median Case Duration	68	55	167
Bond Amount	20655	26914	88423
Defendant Demographics			
Def. Black	0.63	0.78	0.81
Def. Hispanic	0.17	0.1	0.1
Def. White	0.19	0.12	0.09
Def. Male	0.81	0.82	0.92
Def Age	35.85	37	32.62
Case Characteristics			
Charge - Felony	1	1	1
Charge - Felony Violent	0.03	0.02	0.15
Charge - Felony Drug Poss.	0.63	0.66	0.35
Charge - Felony Drug Deliv.	0.11	0.19	0.18
Charge - Felony Property	0.15	0.15	0.17
Charge - Felony Weapon	0.07	0.01	0.19
Charge - Misdemeannaor	0.22	0.18	0.24
Charge - Misd. Property	0.02	0.01	0.02
Charge - Traffic	0.16	0.1	0.08
Charge - Other	0.05	0.05	0.04
Case History (since 2000)			
Past Cases	5.38	8.01	9.09
Case within Year	0.29	0.42	0.54
Past Failure to Appear	1.18	1.85	2.03
Past Felonies	1.61	2.69	3
Past Felonies - Violent	0.07	0.09	0.22
Past Guilty	0.52	1.31	2.28
Past Guilty Felonies	0.19	0.57	1.07
Past Guilty Felonies - Violent	0.01	0.03	0.1
N Obs	13039	19523	18808
Share of Obs	0.25	0.38	0.37

Note: Table displays summary statistics for the main felony sample by pretrial treatment for defendants with one observation per defendant and arrest, meaning if there is a felony and a misdemeanor case against a defendant both sets of charges are aggregated to one observation. Variables beginning with 'Charge' are binary variables indicating any charge of a specific type.

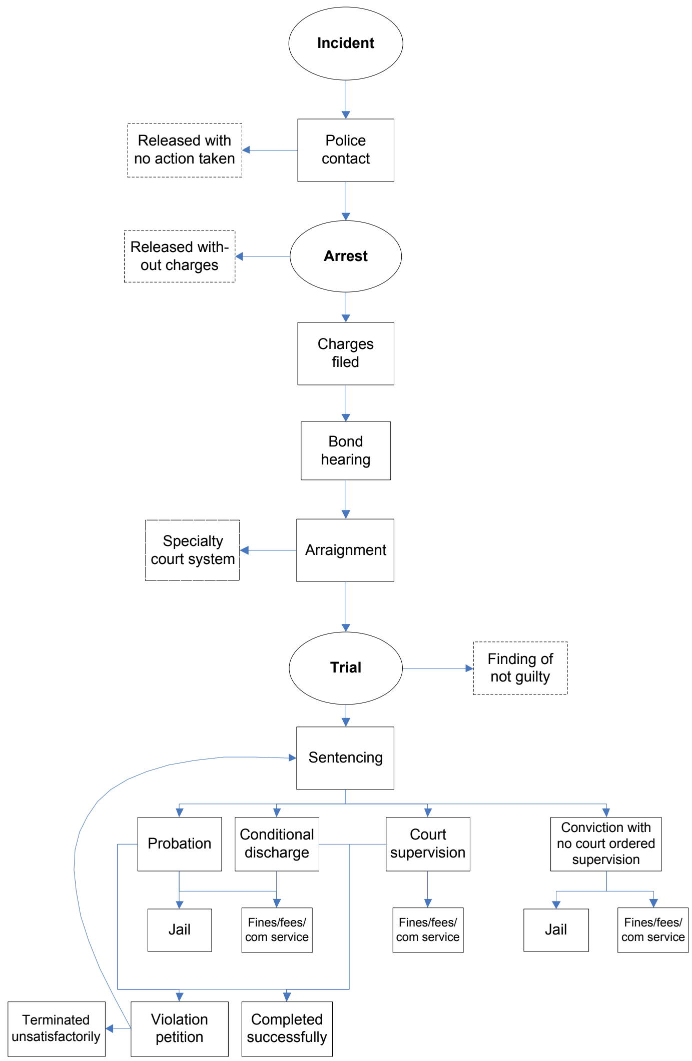
Table A.1: Summary Statistics for Branch 1 Cases by Pretrial Treatment for Main Felony Sample

	Release (1)	EM (2)	Detain (3)
Failure to Appear	0.1220	0.0732	0.0292
Any Guilty Felony Charge	0.345	0.453	0.701
New Case Pretrial	0.171	0.187	0.109
New Case Post-Trial within 4 Years	1.09	1.56	1.58
New Felony Case Post-Trial within 4 Years	0.458	0.656	0.634
New Violent Felony Case Post-Trial within 4 Years	0.0328	0.0405	0.0594

Note: Table displays summary statistics of outcomes for the main felony sample by pretrial treatment for defendants with one observation per defendant and arrest, meaning if there is a felony and a misdemeanor case against a defendant both sets of charges are aggregated to one observation.

Table A.2: Summary Statistics of Outcomes for Main Felony Sample

Figure 2
Flowchart of the general adult misdemeanor criminal process in Illinois



3

Figure A.2: Misdemeanor Case Flow Chart

Note: Figure displays sequence of events for misdemeanor cases within Cook County — though document is meant for the entire Illinois criminal justice system more broadly. Source: [290], page 3.

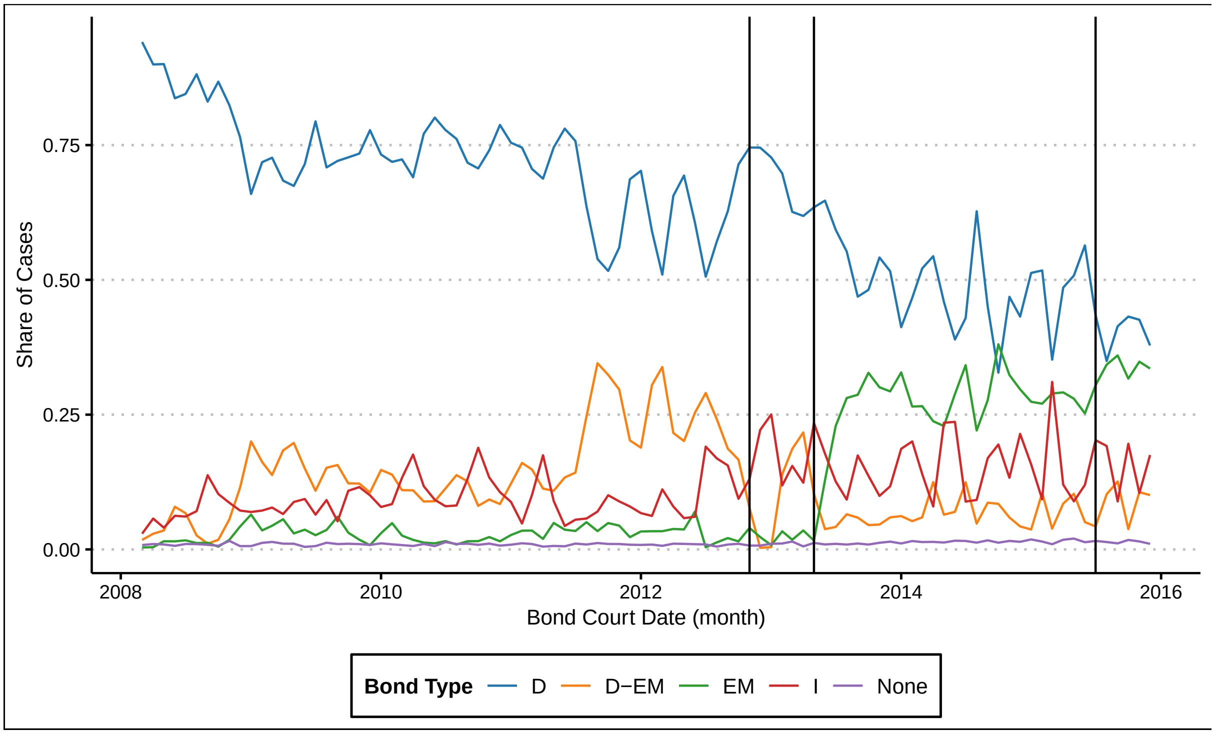


Figure A.3: Bond Time Trends in Cook County Court

Note: Figure displays the composition of bond types within the sample between 2008 and 2015 aggregated by year-month of bond date. D-EM bond refers to D-bonds coupled with EM release, and EM refers to I-bonds coupled with EM (IEM). The first vertical line denotes November 2012, when D-EM bonds stopped being issued temporarily, and the second vertical line denotes the introduction of IEM bonds in June 2013.

Outcome	All	Release (S=1)	EM (S=2)	Detain (S=3)
Fail to Appear	0.069	0.122	0.073	0.029
New Case Pretrial	0.154	0.171	0.187	0.109
New Case Pretrial - Serious	0.058	0.061	0.063	0.051
New Case Pretrial - Low-Level	0.122	0.147	0.136	0.091
New Case Pretrial - Violation/ Escape	0.021	0.013	0.042	0.005
Any Guilty Felony Charge	0.517	0.345	0.453	0.701
Sentenced to Incarceration	0.363	0.139	0.284	0.601
Total New Cases Post Trial within 4 Years	1.450	1.090	1.560	1.580
Total New Cases Post Trial within 4 Years - Felony	0.598	0.458	0.657	0.634
Total New Case Cost Over 4 Years	58.100	44.900	49.600	75.900
Total New Case Cost Over 4 Years (Low Murder)	24.200	19.900	25.000	26.300
Total New Case Pretrial Cost	7.190	8.740	5.580	7.780
Total New Case Pretrial Cost (Low Murder)	3.580	4.190	3.680	3.050
Total New Case Post-Trial Over 3 Years Cost	38.700	28.300	35.400	49.400
Total New Case Post-Trial Over 3 Years Cost (Low Murder)	15.500	12.400	16.400	16.700

Note: Table displays summary statistics of all outcomes for the main felony sample for all treatments and each pre-trial treatment for defendants.

Table A.3: Outcomes Means for Main Felony Sample

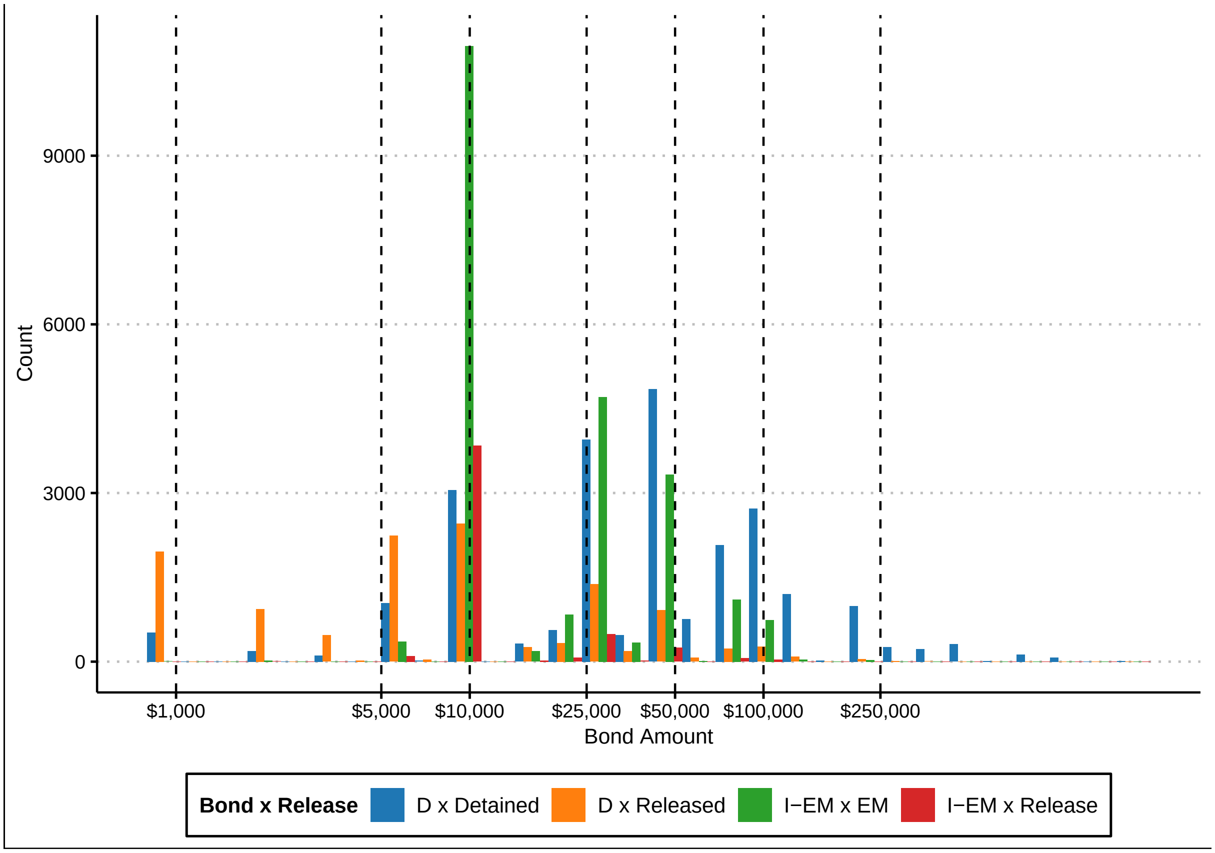


Figure A.4: Distribution of Bond Amounts

Note: Figure displays histogram of bond amounts (price of release being 10% of the bond amount) for IEM and D bonds for each pretrial treatment during the sample period. X-axis is scaled by the log of bond amount.

	Release		Detain
Hansen Over-Identification			
J-Statistic	445.088		446.462
P-value	<0.001		<0.001
Montiel Olea - Pflueger Weak Instrument			
Eff. Fstat	35.793		54.304
2SLS Critical Value (5% of Worst Case Bias)	22.593		22.692
Anderson-Rubin Weak Instrument			
F-statistic		5.2	
P-Value		<0.001	

Note: Table displays results for additional instrument tests. Eff. Fstat refers to the effective F-statistic from the [124] weak instruments test.

Table A.4: Additional Instrument Tests

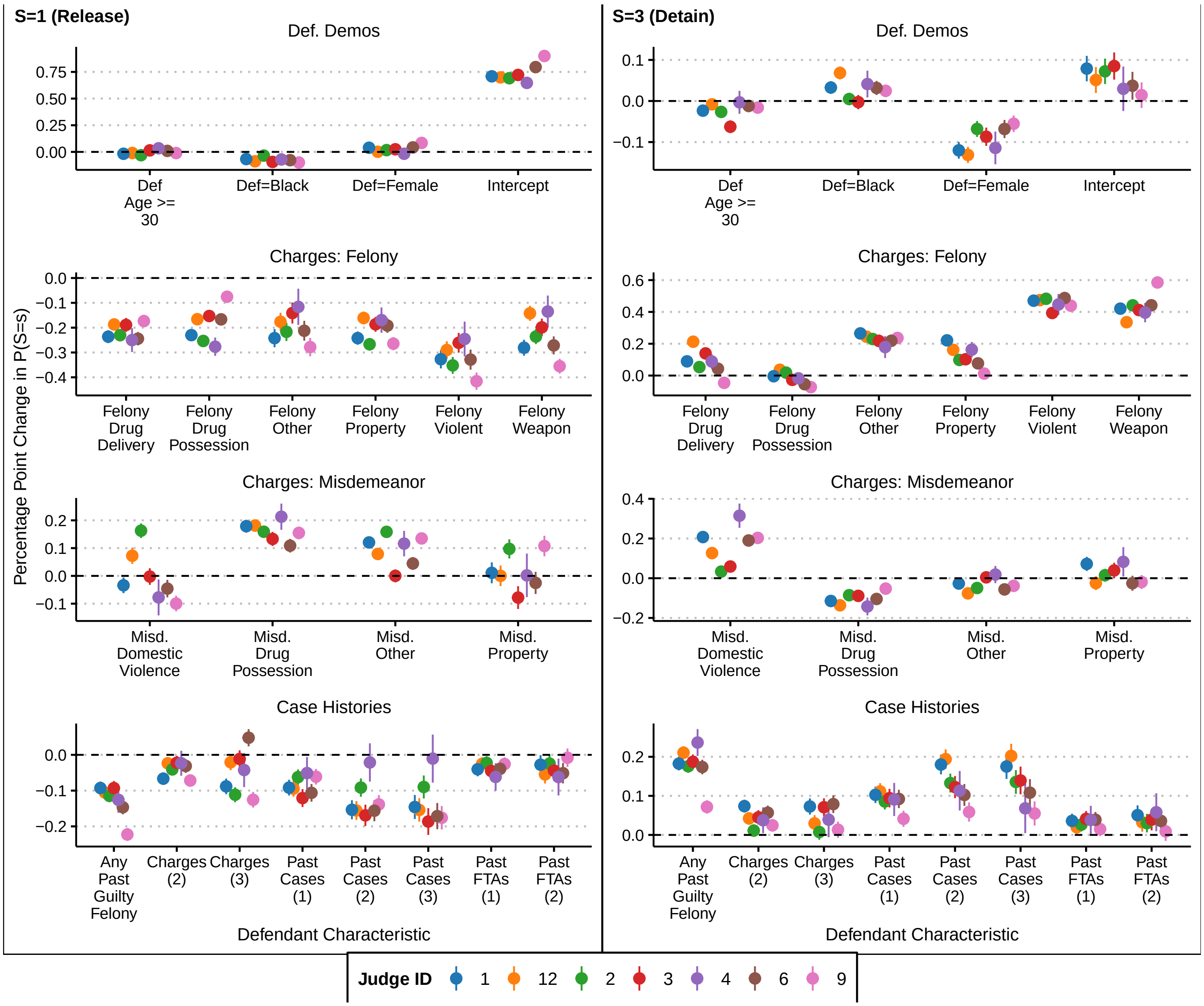


Figure A.5: Judge Preferences over Defendant Observables

Note: Figure displays OLS estimates of judge-specific coefficients (points, with 95% confidence intervals) for each defendant characteristic (binary variables) recovered by estimating an LPM where pretrial treatment being either Release (S=1) or Detention (S=3) is the outcome, including month and day of week fixed effects (equation (4)). Intercept is the baseline probability of the treatment by the judge. Case history variables involve bins: Charges (2) refers to between 3 and 10 charges total, and Charges (3) refers to more than 10 charges total; Past Cases (1) has between 1 and 4 past cases, (2) has between 5 and 12, and (3) has more than 12; Past FTAs (1) has either 1 or 2 FTAs in past cases, and (2) has more than 2.

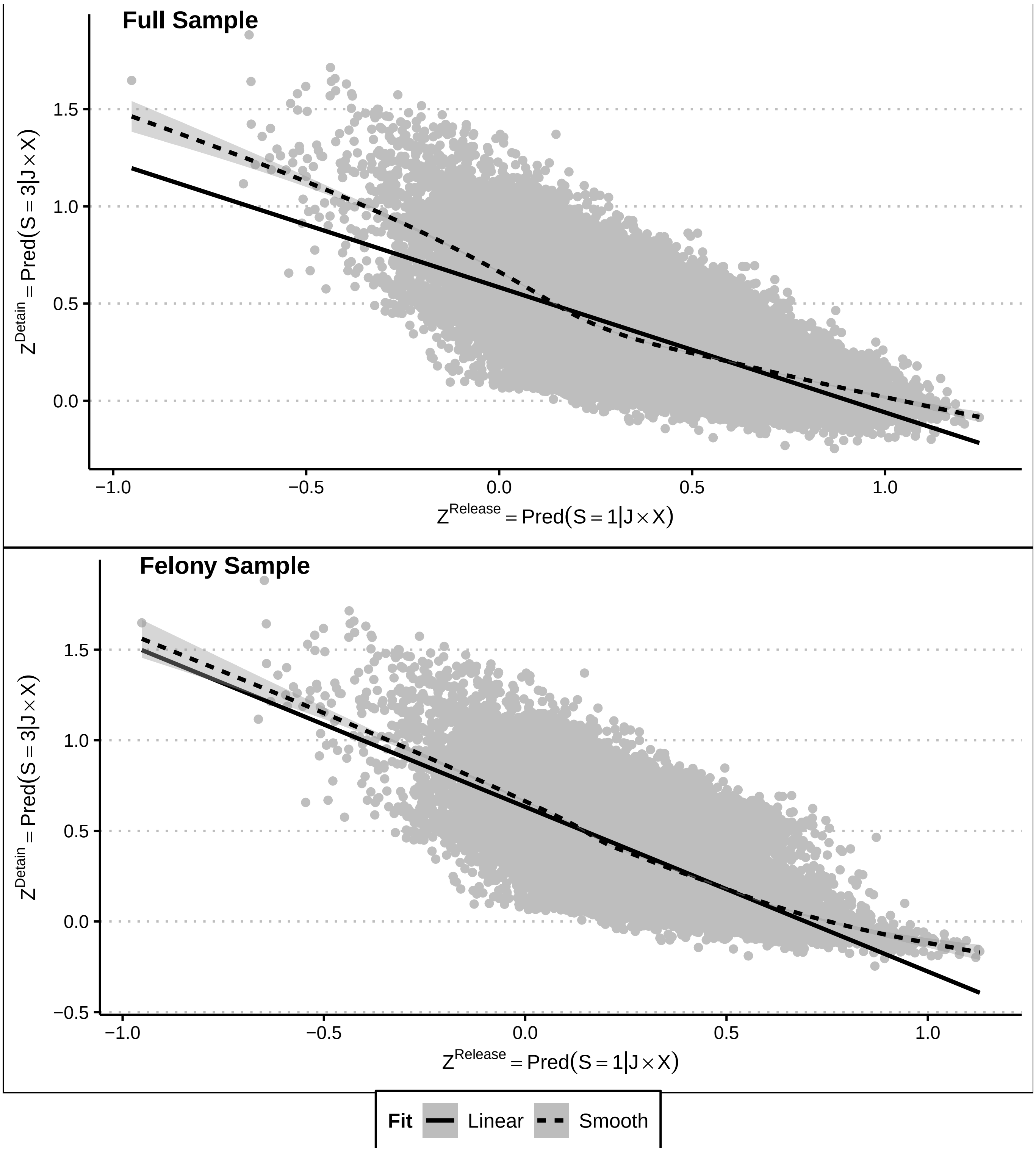


Figure A.6: Testing for Linearity Between Predicted Treatments

Note: Figure displays scatter plot of observations by values of instruments for Release and Detention as well as their relationship with a linear fit (solid line) and nonlinear fit (dashed line), for both full (top) and main felony (bottom) samples. Nonlinear fit is computed using ‘loess’ local polynomial regression.

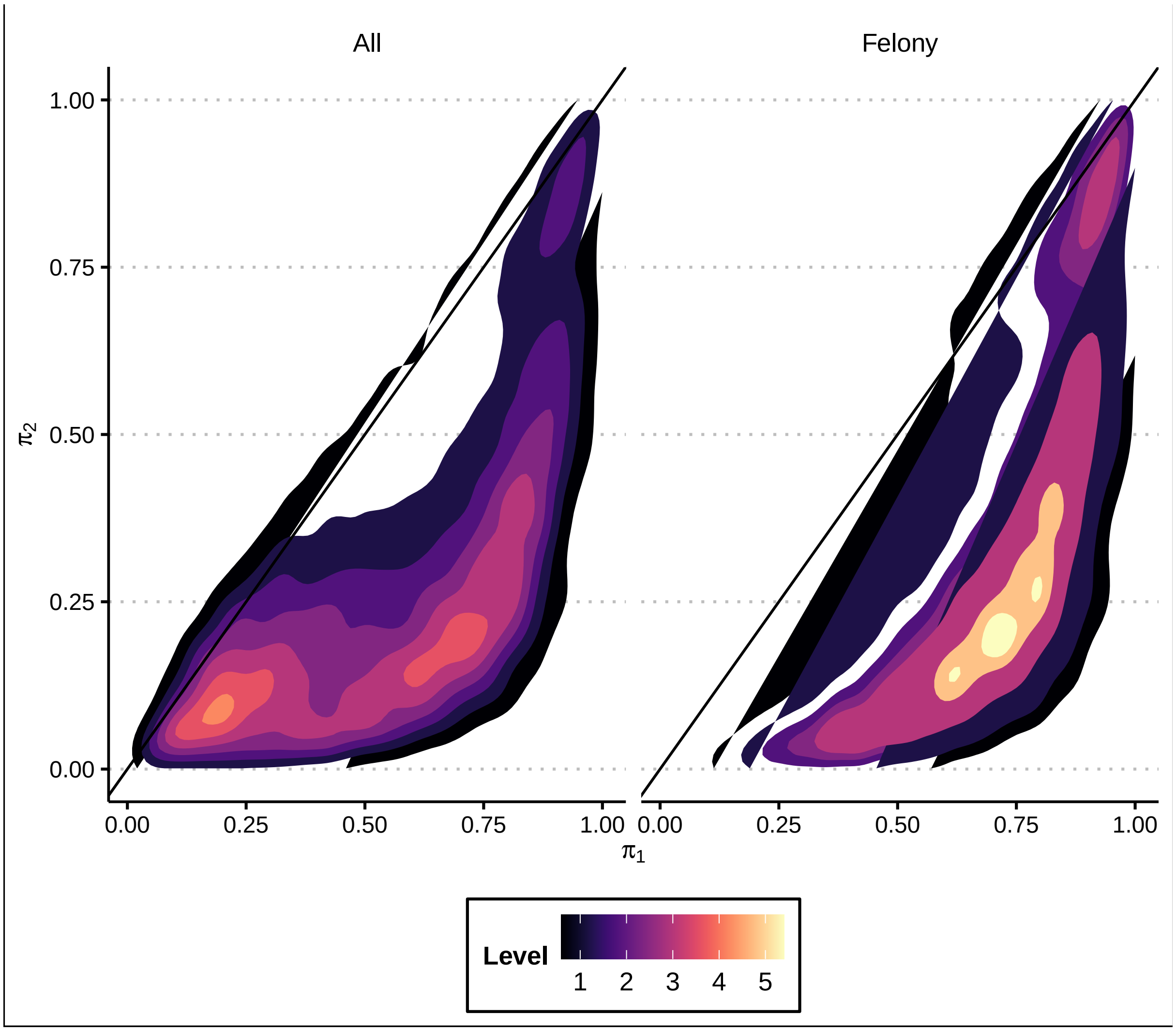


Figure A.7: Density of π_1 and π_2 by Sample

Note: Figure displays density of fitted values for π_1 and π_2 for the full and main felony samples.

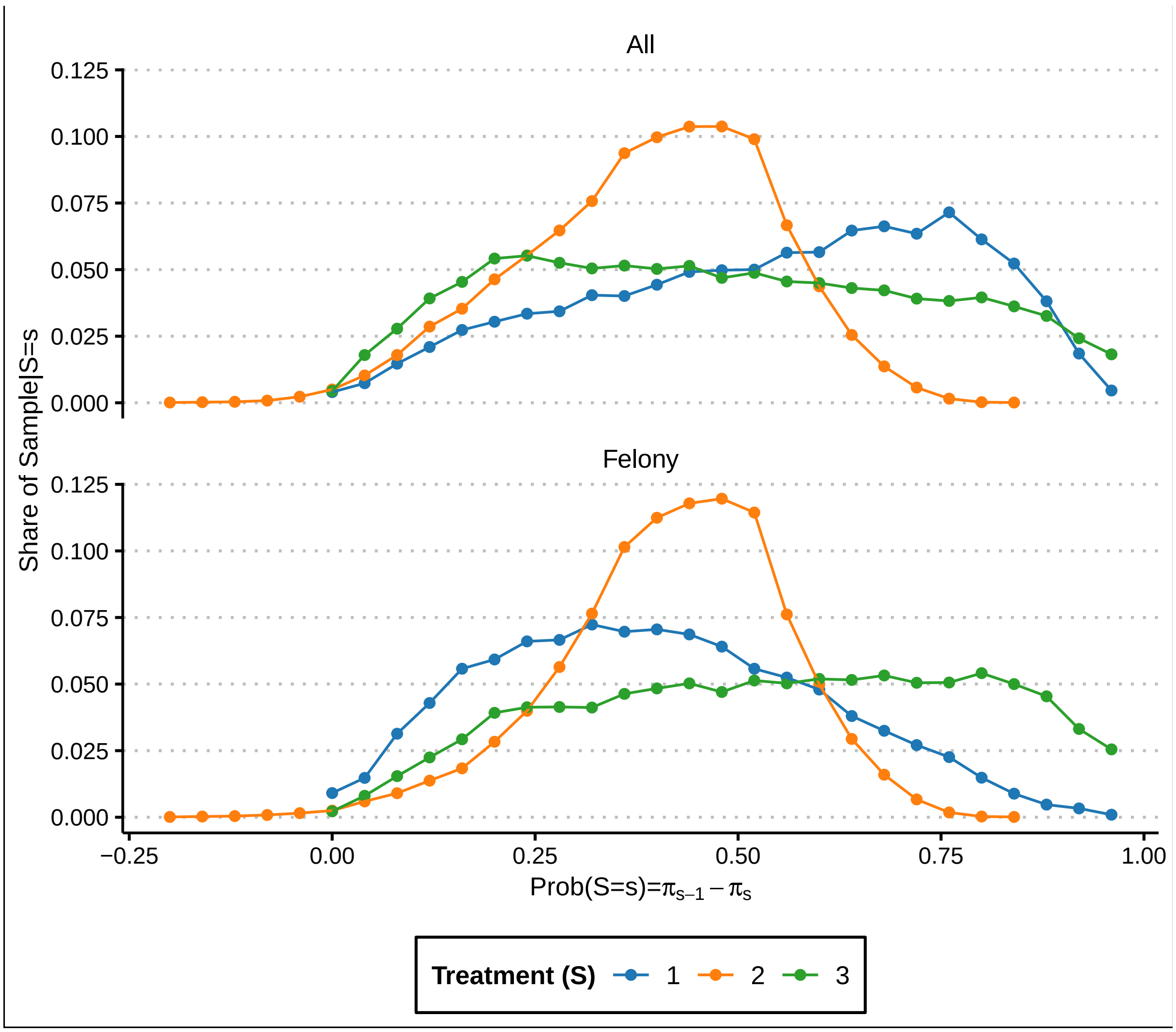


Figure A.8: Density of $\Pr(S = s) = \pi_{s-1} - \pi_s$

Note: Figure displays densities, within each treatment, of the predicted likelihood the defendant is assigned to that treatment for both full and main felony samples.

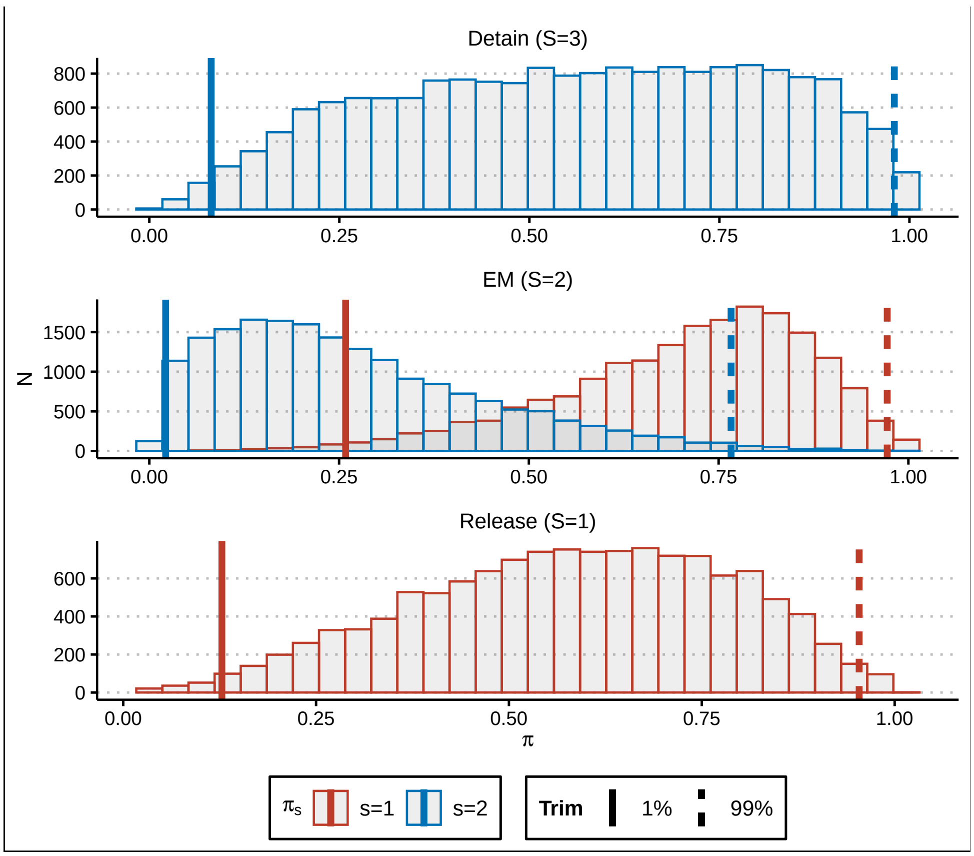


Figure A.9: Common Support Across Treatment Levels

Note: Figure displays histograms of relevant π_s and π_{s-1} values for each treatment level (s). Trims indicate the 1st percentile and 99th percentile of the treatment-specific sample for each relevant π value.

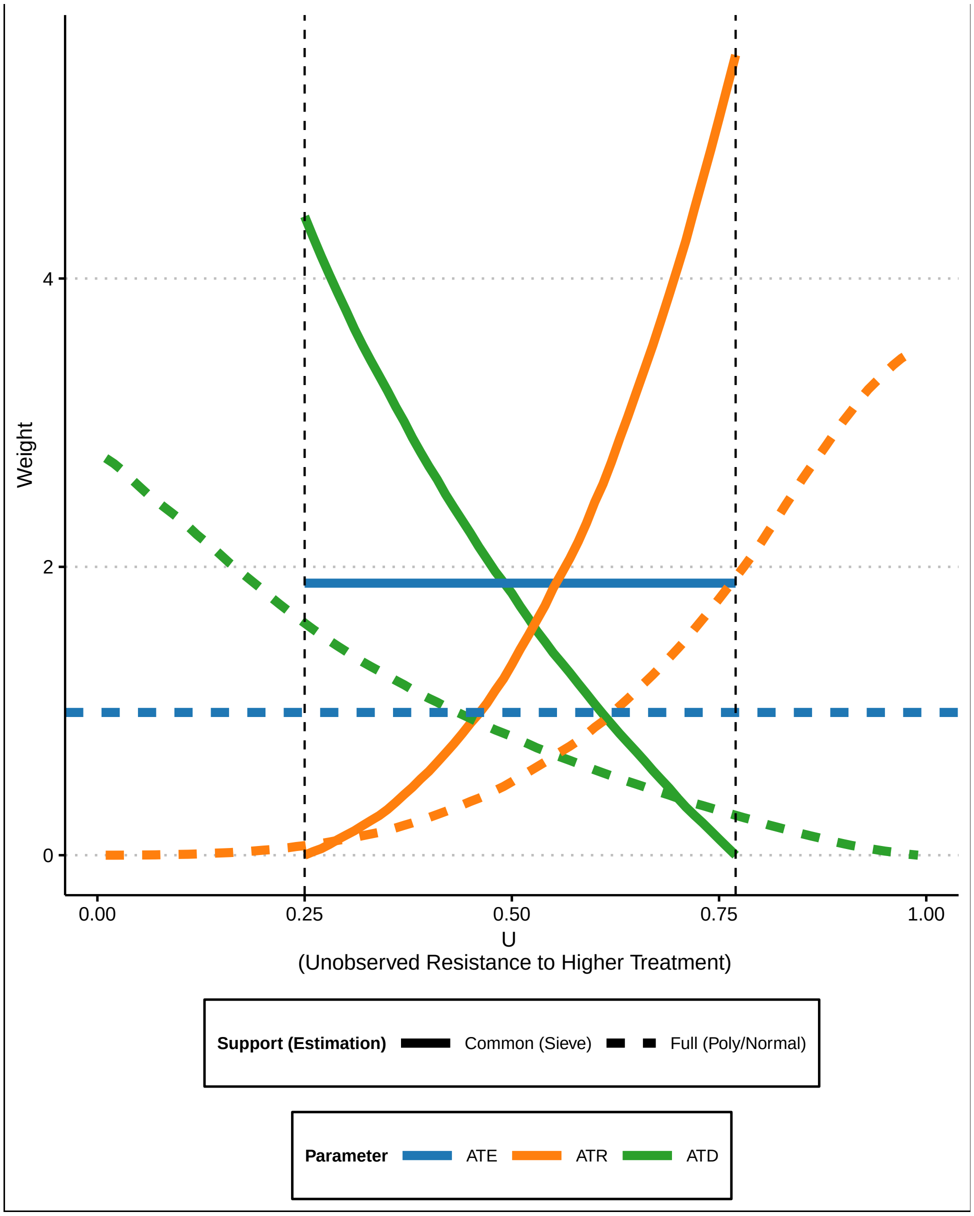


Figure A.10: Treatment Effect Weights

Note: Figure displays the weights used to sum over unobserved resistance to treatment ($U = \pi_1, \pi_2$ depending on the treatment margin) to compute each treatment effect. ATR(D) means average treatment effect within common support on the released (detained) defendants and thus applies a higher weight to higher (lower) resistance to higher treatment. The ATE applies equal weights as u is distributed uniformly.

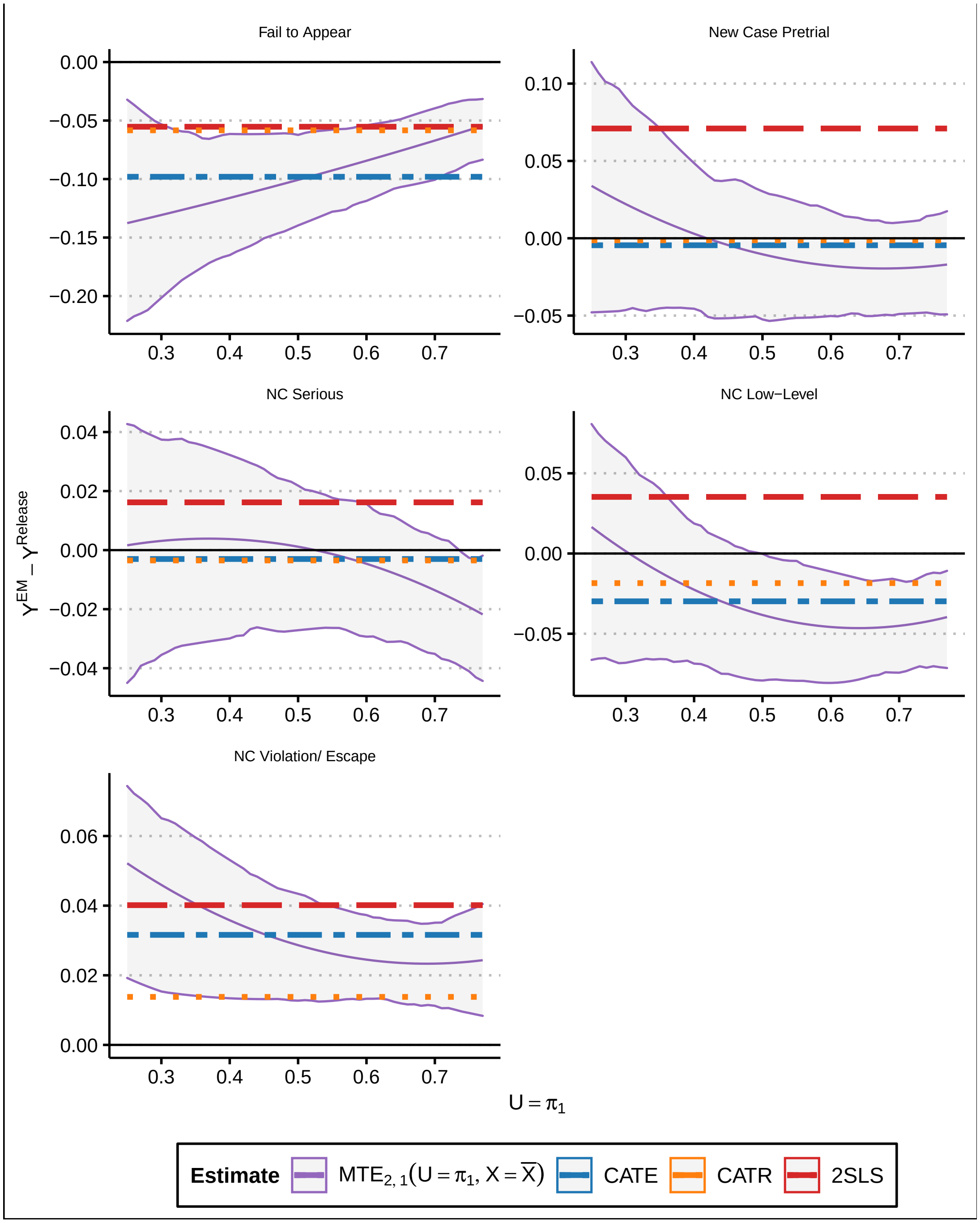


Figure A.11: EM vs. Release MTEs for Pretrial Misconduct

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Release for main felony sample cases. Outcomes consist of variables relating to pretrial misconduct. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATR, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be released. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

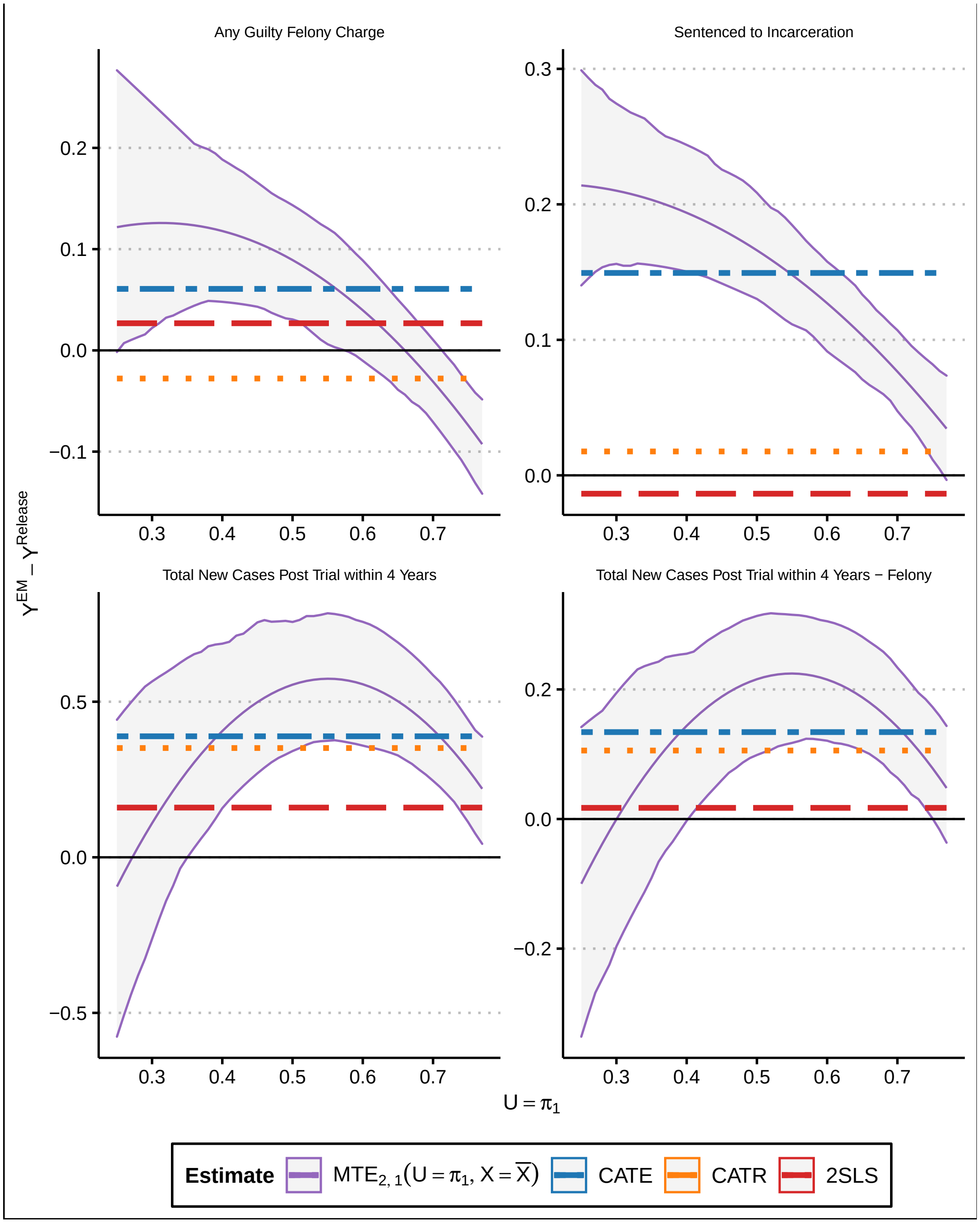


Figure A.12: EM vs. Release MTEs for Case and Post-Trial Outcomes

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Release for main felony sample cases. Outcomes consist of variables relating to case outcomes and post-trial recidivism. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATR, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be released. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

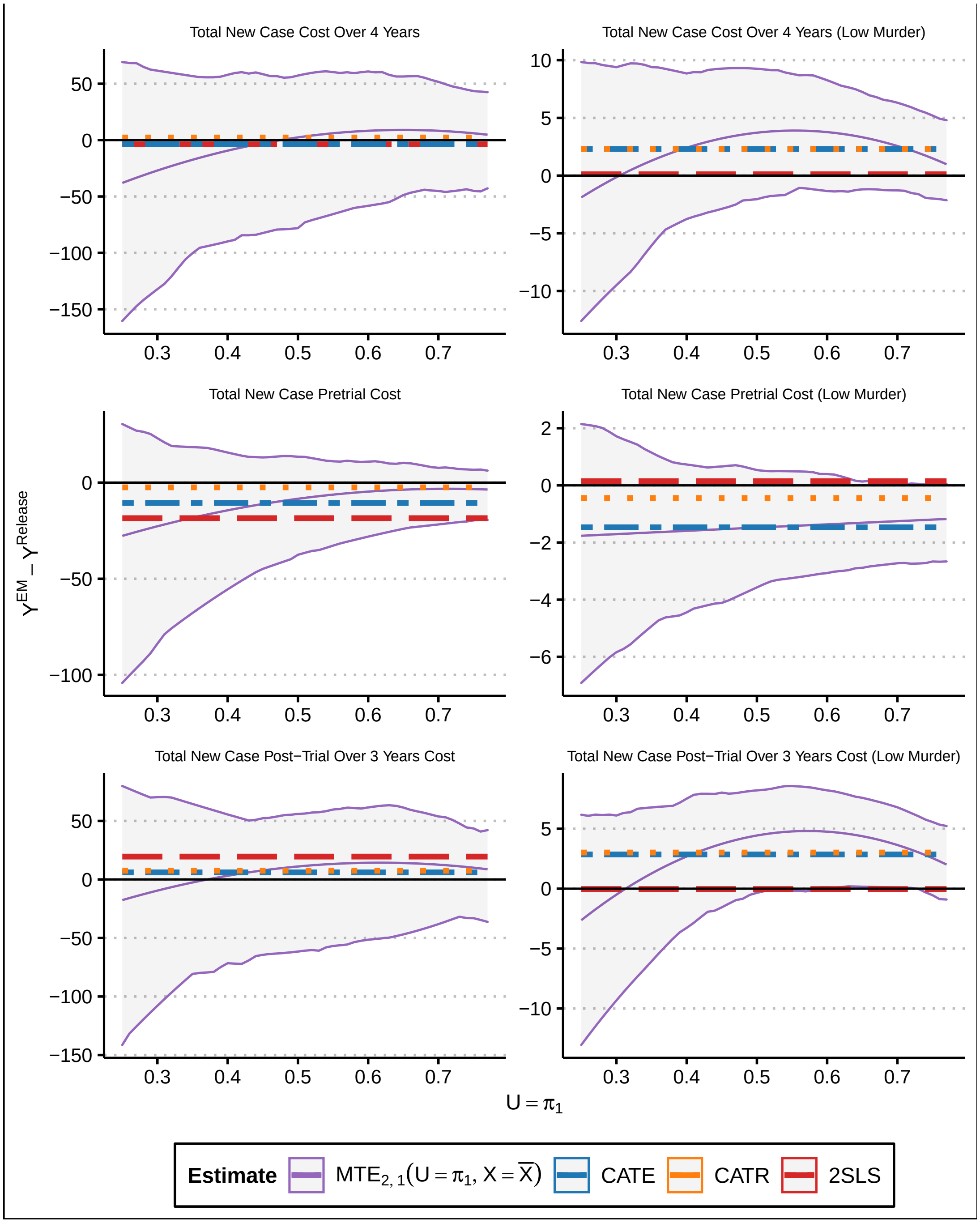


Figure A.13: EM vs. Release MTEs for New Case Costs ($1,000)

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Release for main felony sample cases. Outcome consist of total new case costs pre and post-trial using full and low-cost values for murder based on incidence costs from [19]. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATR, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be released. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

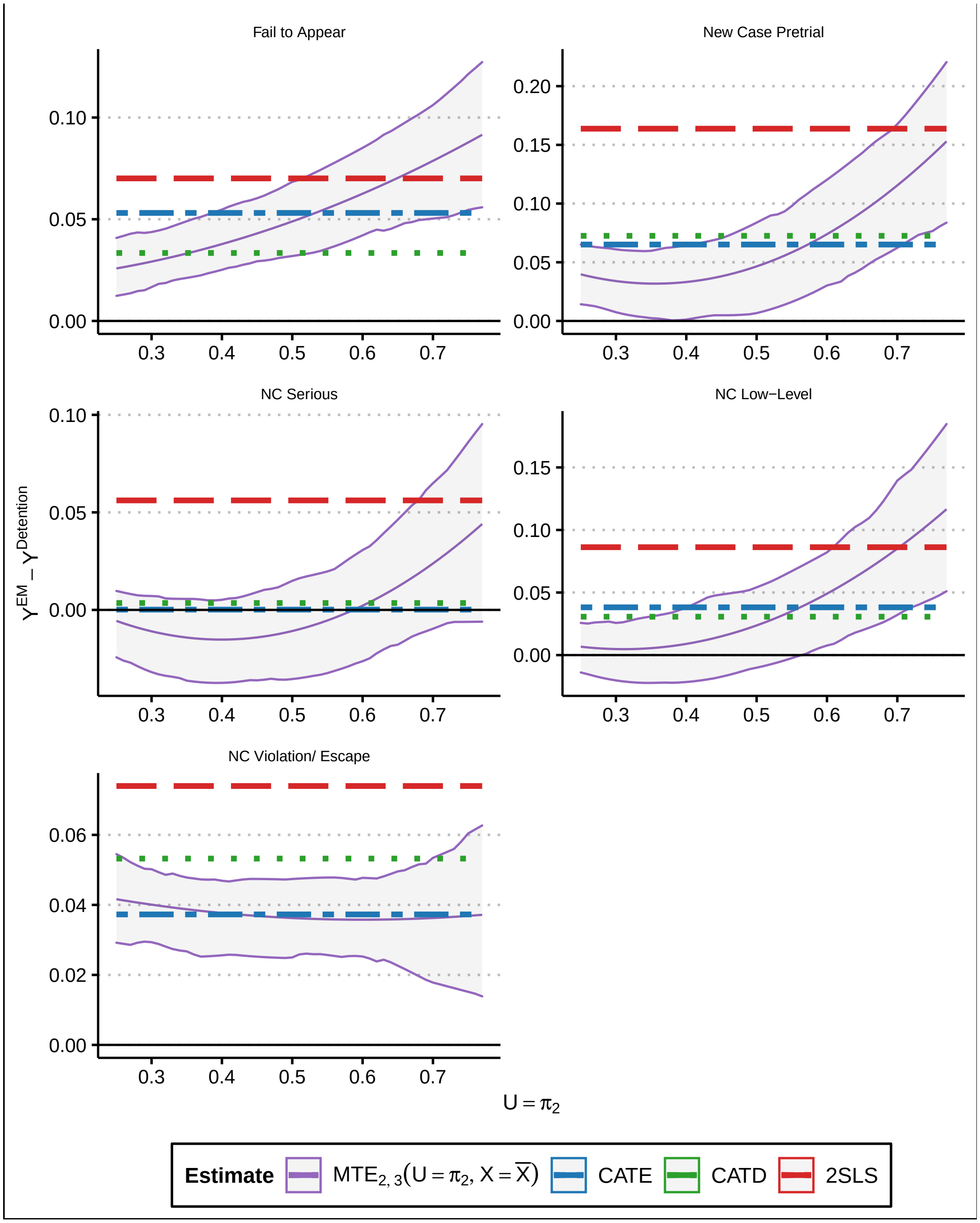


Figure A.14: EM vs. Detention MTEs for Pretrial Misconduct

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Detention for main felony sample cases. Outcomes consist of variables relating to pretrial misconduct. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATD, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be detained. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

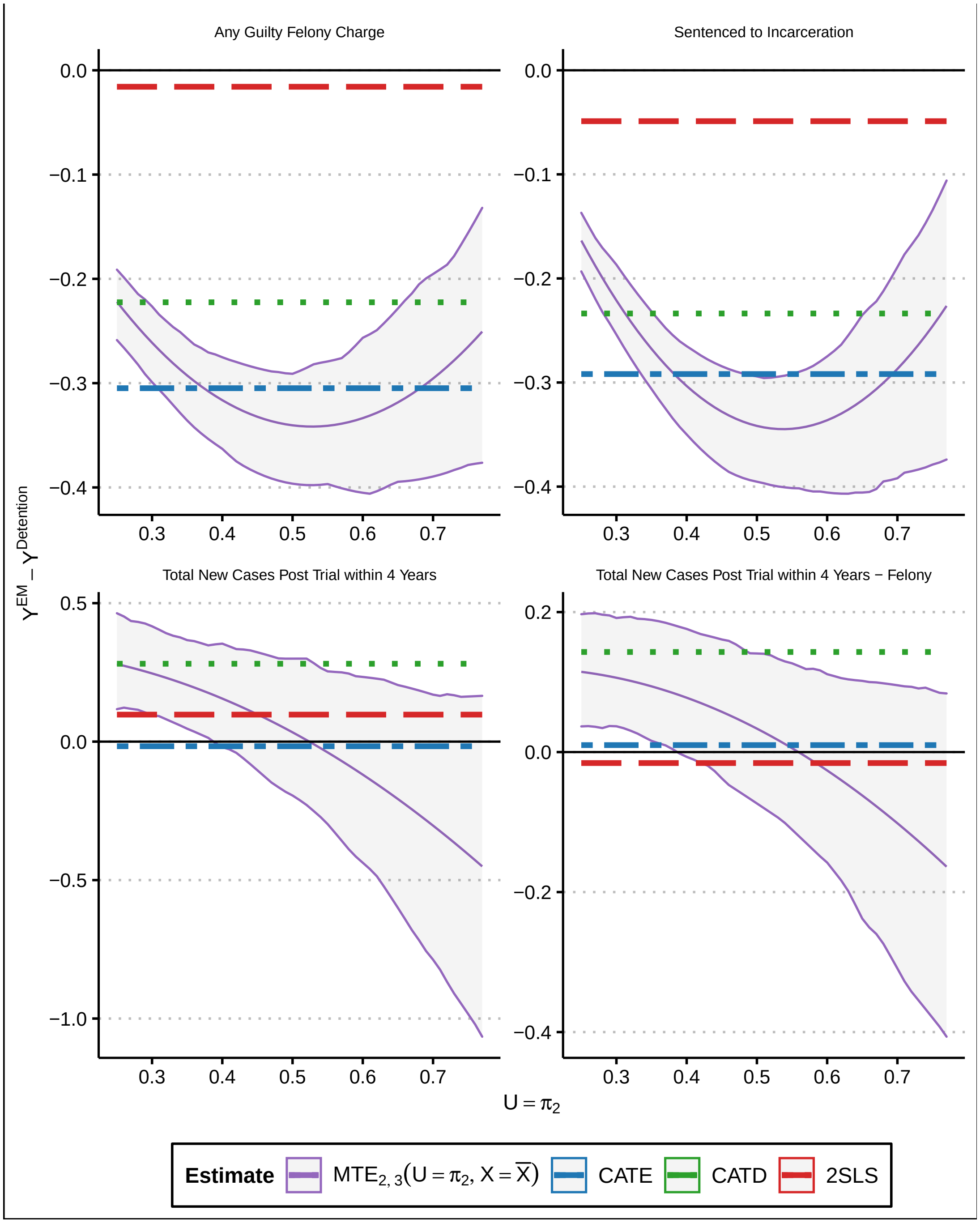


Figure A.15: EM vs. Detention MTEs for Case and Post-Trial Outcomes

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Detention for main felony sample cases. Outcomes consist of variables relating to case outcomes and post-trial recidivism. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATD, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be detained. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

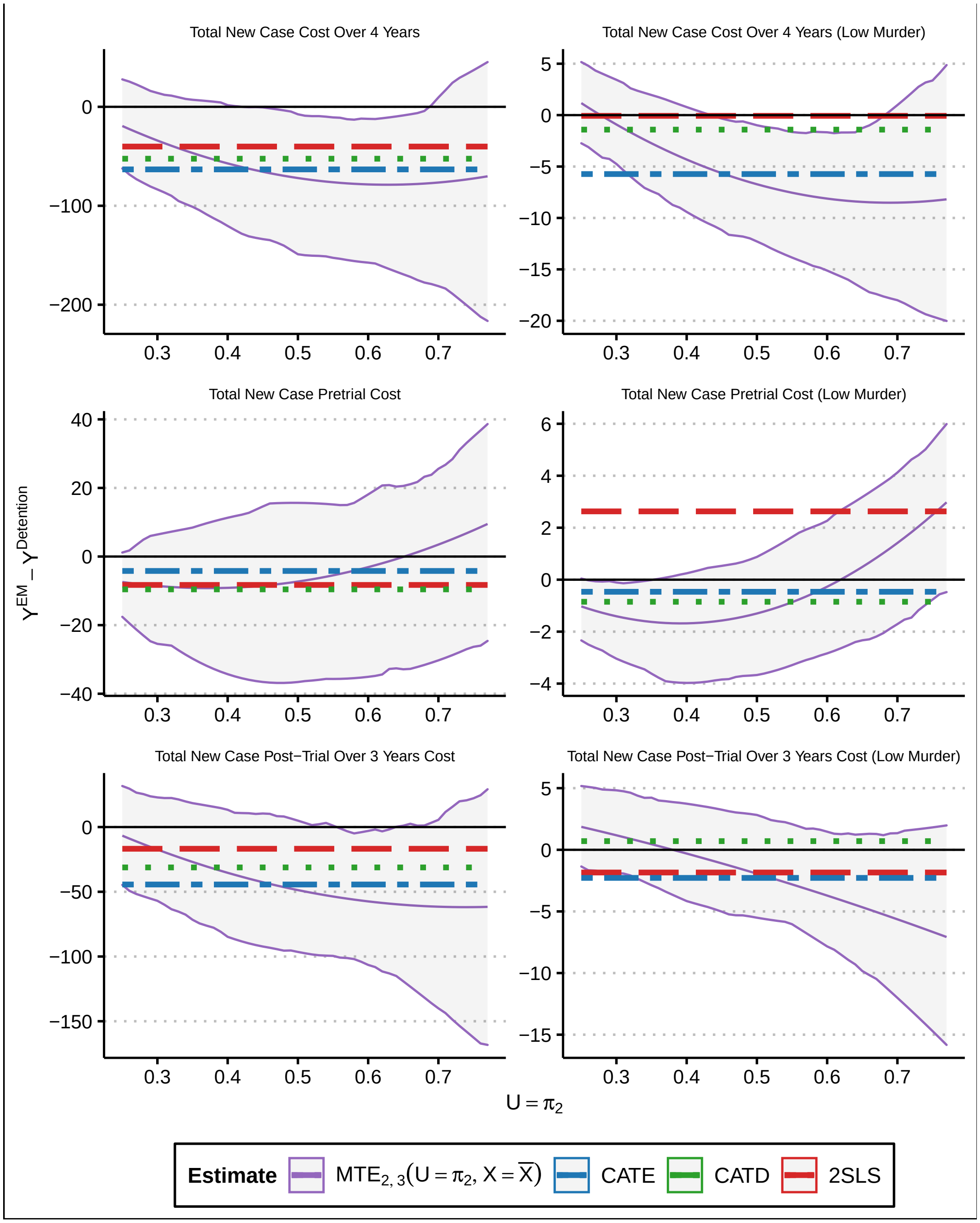


Figure A.16: EM vs. Detention MTEs for New Case Costs ($1,000)

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the effect of EM relative to Detention for main felony sample cases. Outcome consist of total new case costs pre and post-trial using full and low-cost values for murder based on incidence costs from [19]. Horizontal lines denote the corresponding 2SLS estimates and CATE, which averages the MTEs over the common support with equal weights, and the CATD, which averages the MTEs placing higher weights on defendant observably and unobservably more likely to be detained. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

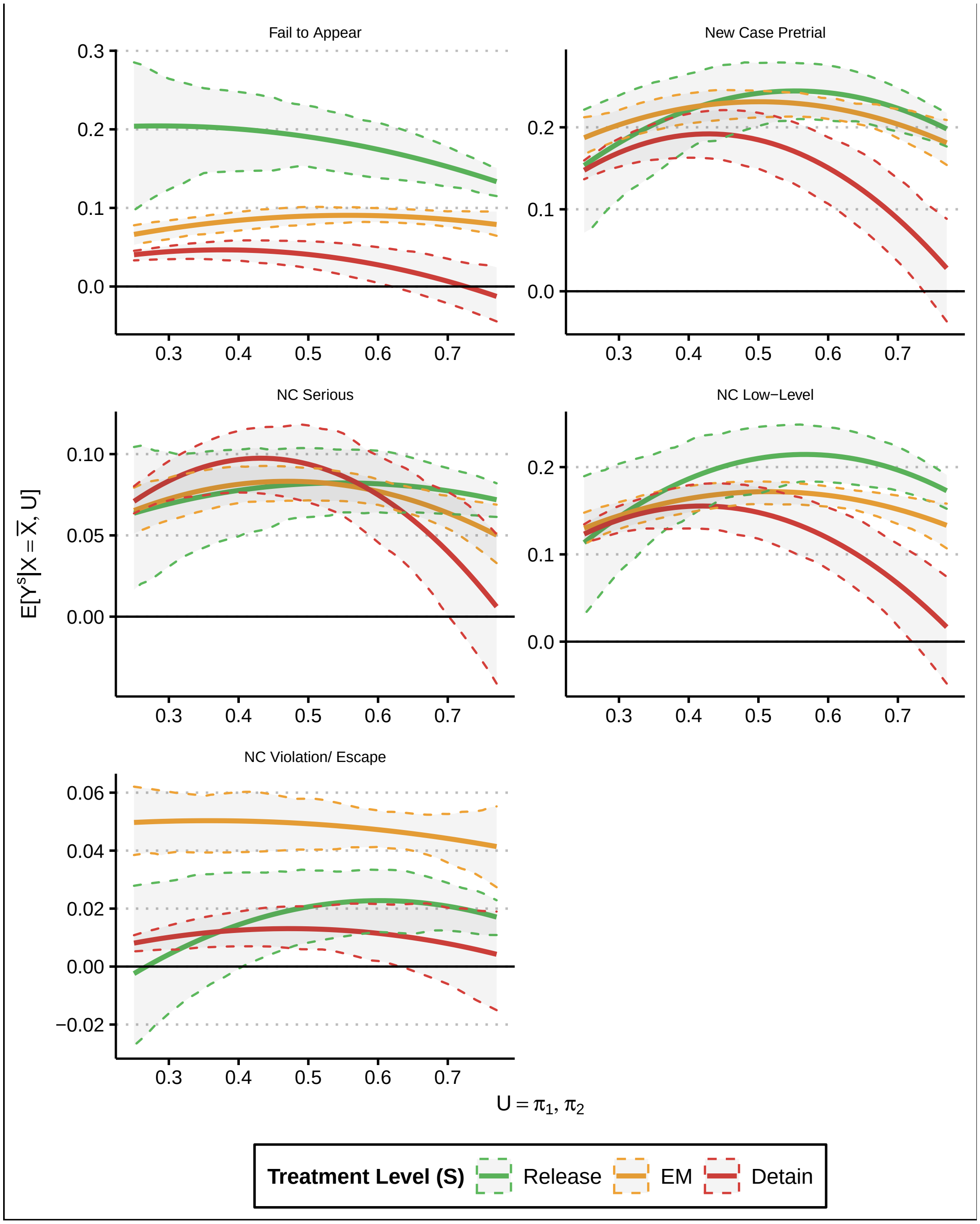


Figure A.17: MTRs for Pretrial Misconduct

Note: Figure displays the marginal treatment response functions (MTR) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the expected response in each treatment level for main felony sample cases for the relevant outcomes. MTRs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

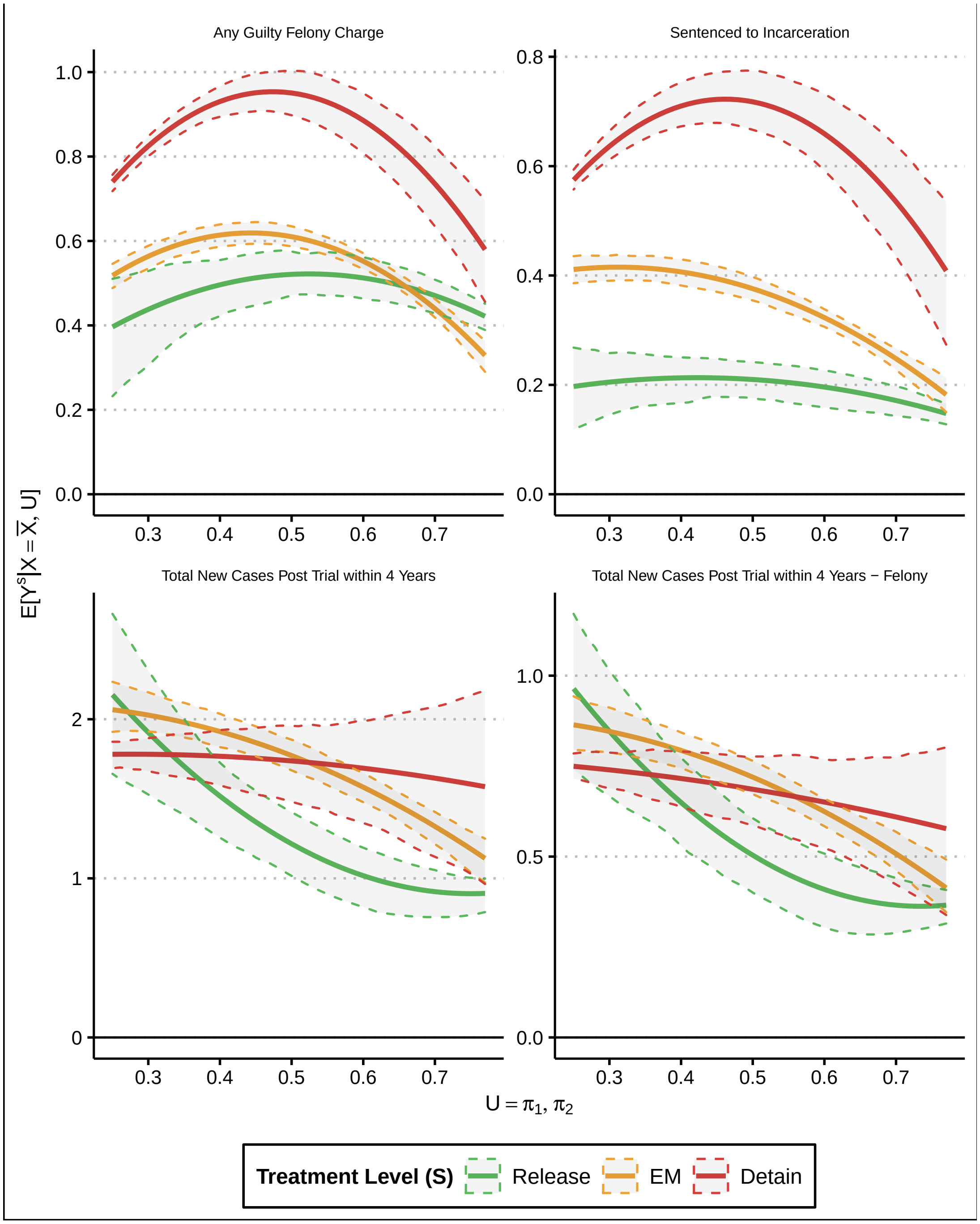


Figure A.18: MTRs for Case and Post-Trial Outcomes

Note: Figure displays the marginal treatment response functions (MTR) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the expected response in each treatment level for main felony sample cases for the relevant outcomes. MTRs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

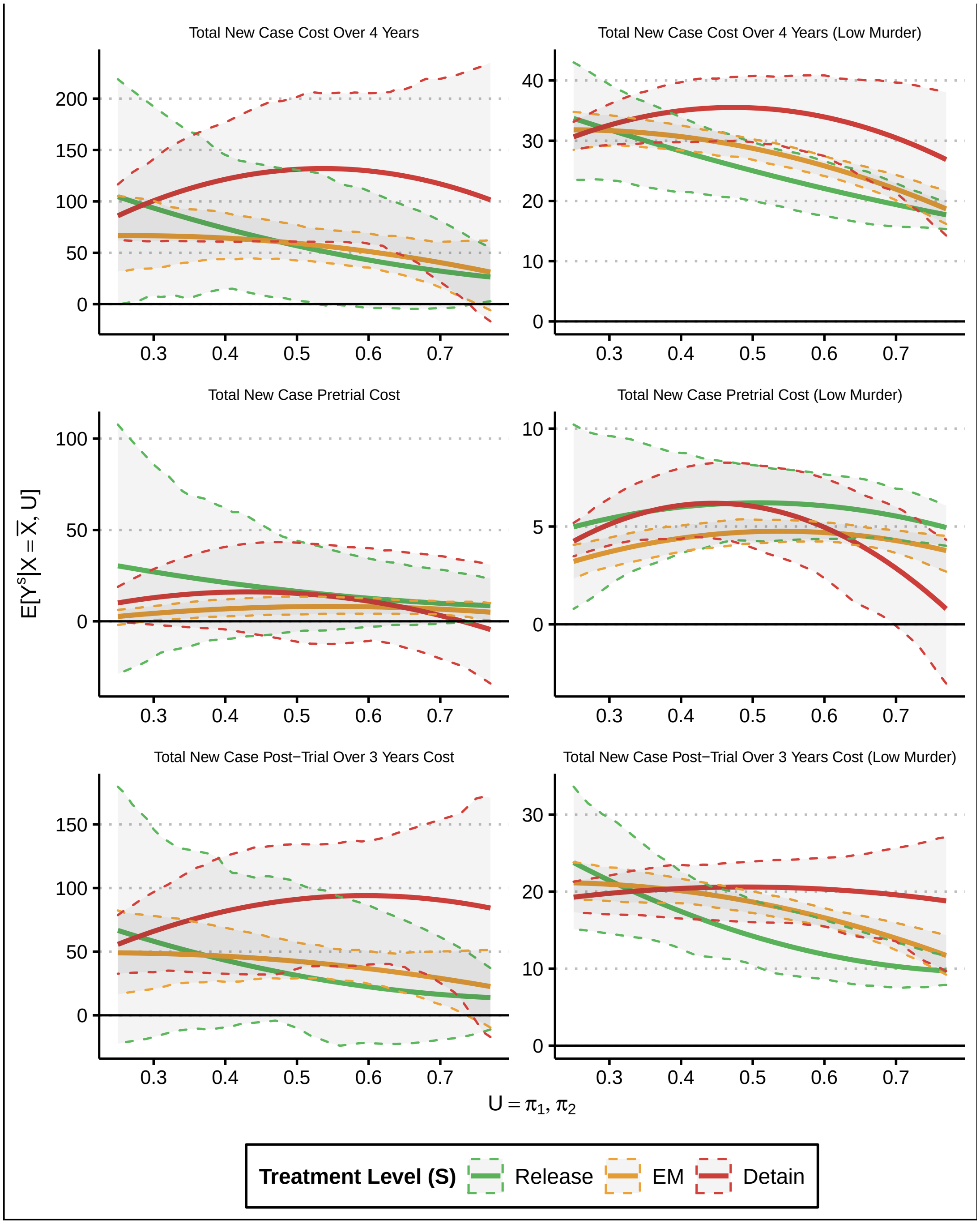


Figure A.19: MTRs for New Case Costs ($1,000)

Note: Figure displays the marginal treatment response functions (MTR) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the expected response in each treatment level for main felony sample cases for the relevant outcomes. MTRs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s). 95% confidence intervals are computed using 200 bootstrap runs.

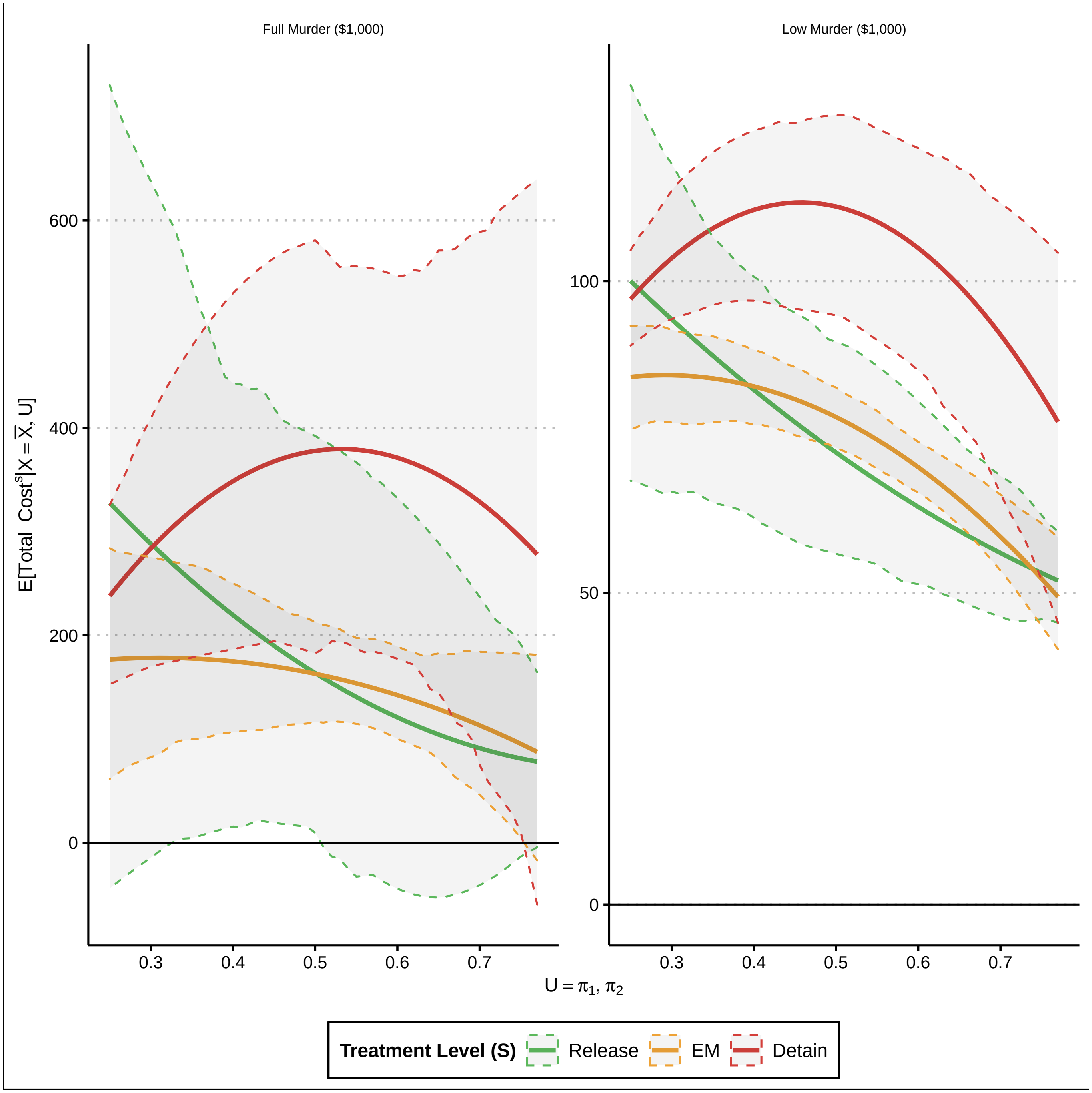


Figure A.20: MTRs for Total Costs ($1,000)

Note: Figure displays the marginal treatment response functions (MTR) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) for the expected total costs (including pre- and post-trial crime, sentencing, failures to appear, and direct costs) in each treatment level for main felony sample cases. MTRs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s) with weights and costs applied as discussed in Section 5.4. 95% confidence intervals are computed using 200 bootstrap runs.

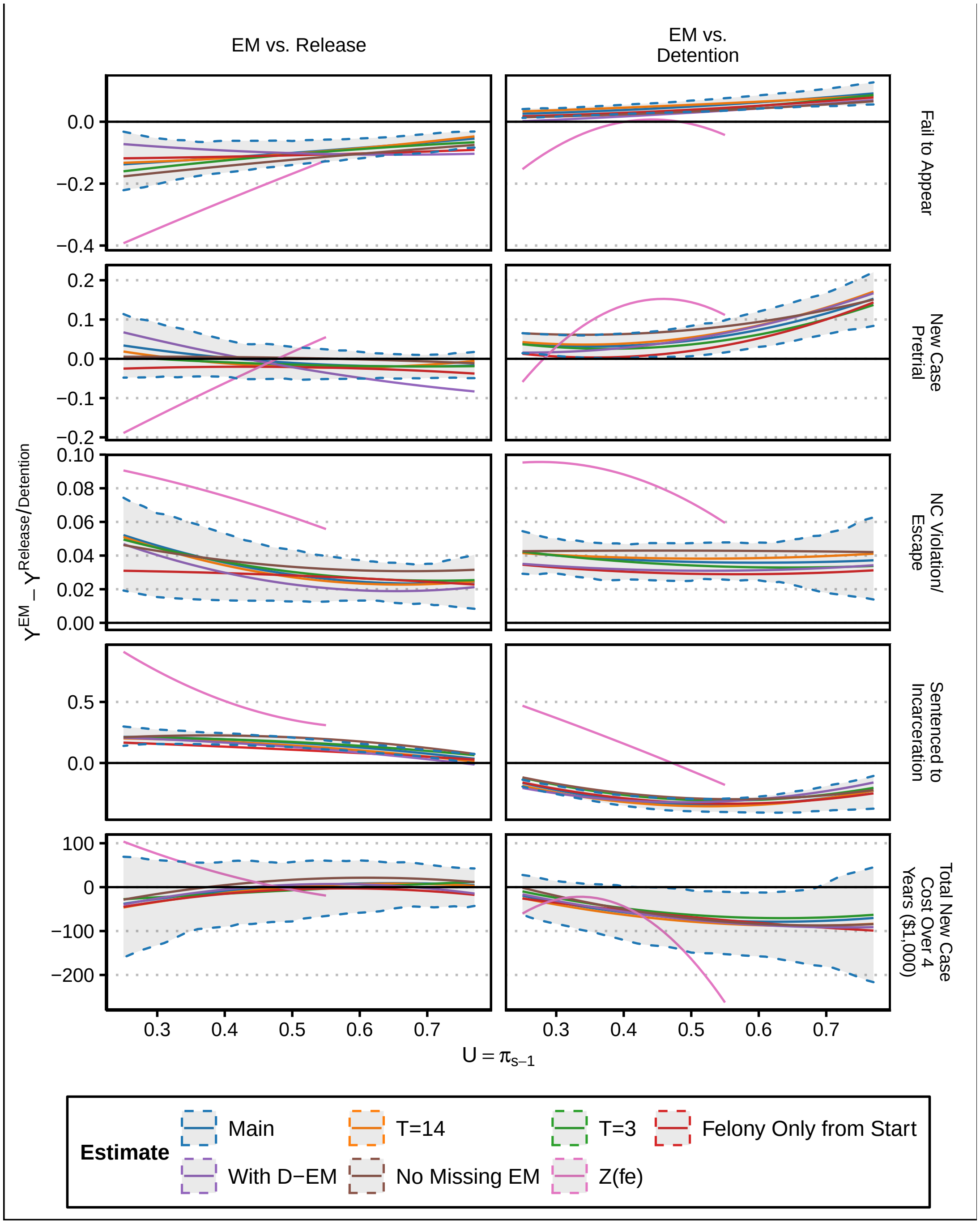


Figure A.21: MTEs for Main Robustness Tests

Note: Figure displays the marginal treatment effects (MTEs) for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right). MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of main estimates are computed using 200 bootstrap runs.

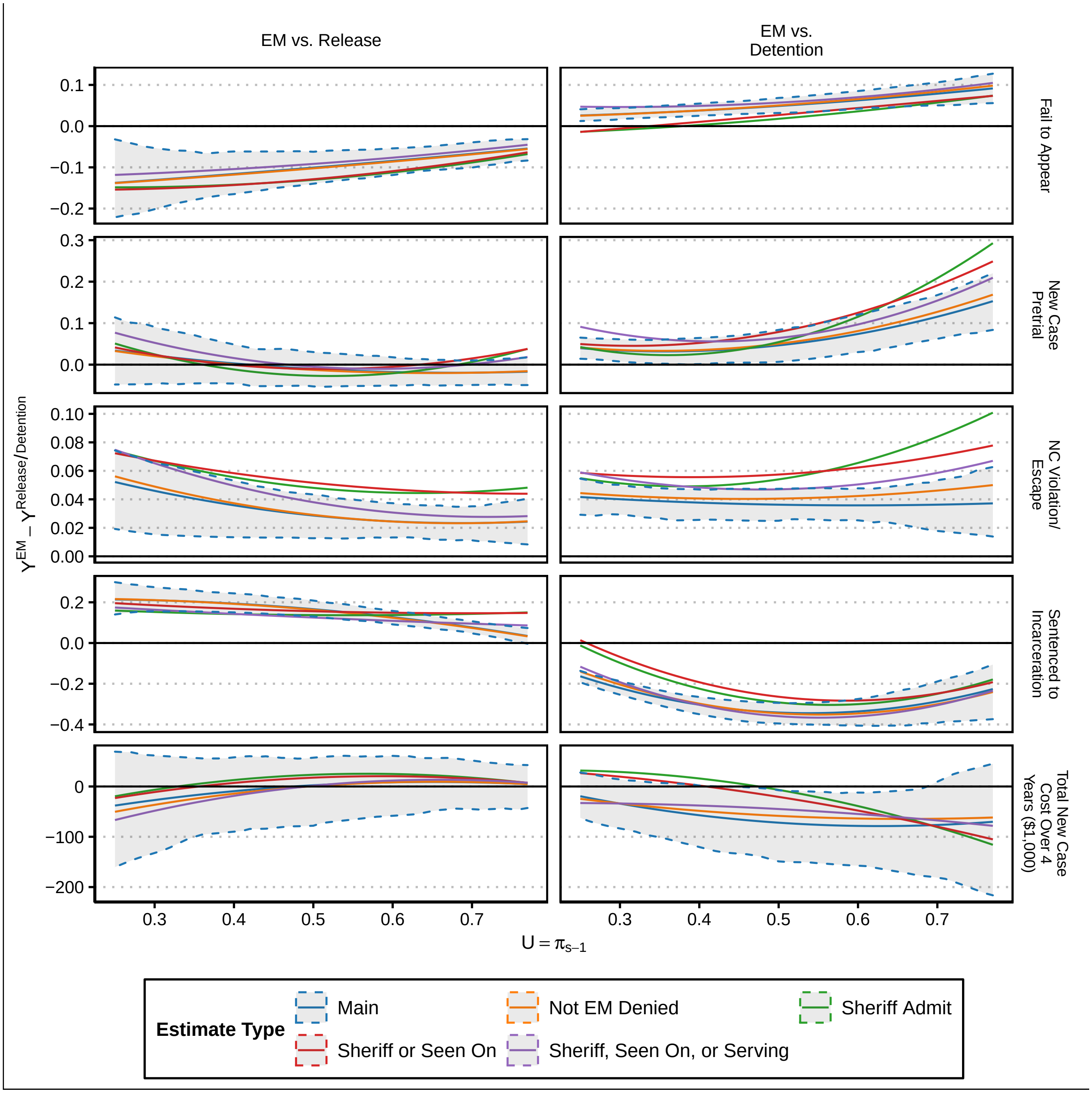


Figure A.22: MTEs for Recoded Treatments using Disposition Codes

Note: Figure displays the marginal treatment effects (MTEs) for various robustness checks for the effect of EM relative to Release (left) and and EM relative to Detention (right) under different re-codings of pretrial treatments using disposition codes to determine if a defendant was assigned to EM. "Sheriff Admit" means the defendant was explicitly noted to have been admitted into the sheriff's EM program; "Sheriff or Seen On" allows for if the defendant was explicitly noted to be on EM; "Sheriff, Seen On, or Serving" allows for if the defendant was explicitly noted to be serving a monitoring program; and "Not EM Denied" includes "Sheriff, Seen On, or Serving" defendants but excludes any defendant explicitly noted to not be admitted to EM (with bail set to stand). MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of main estimates are computed using 200 bootstrap runs.

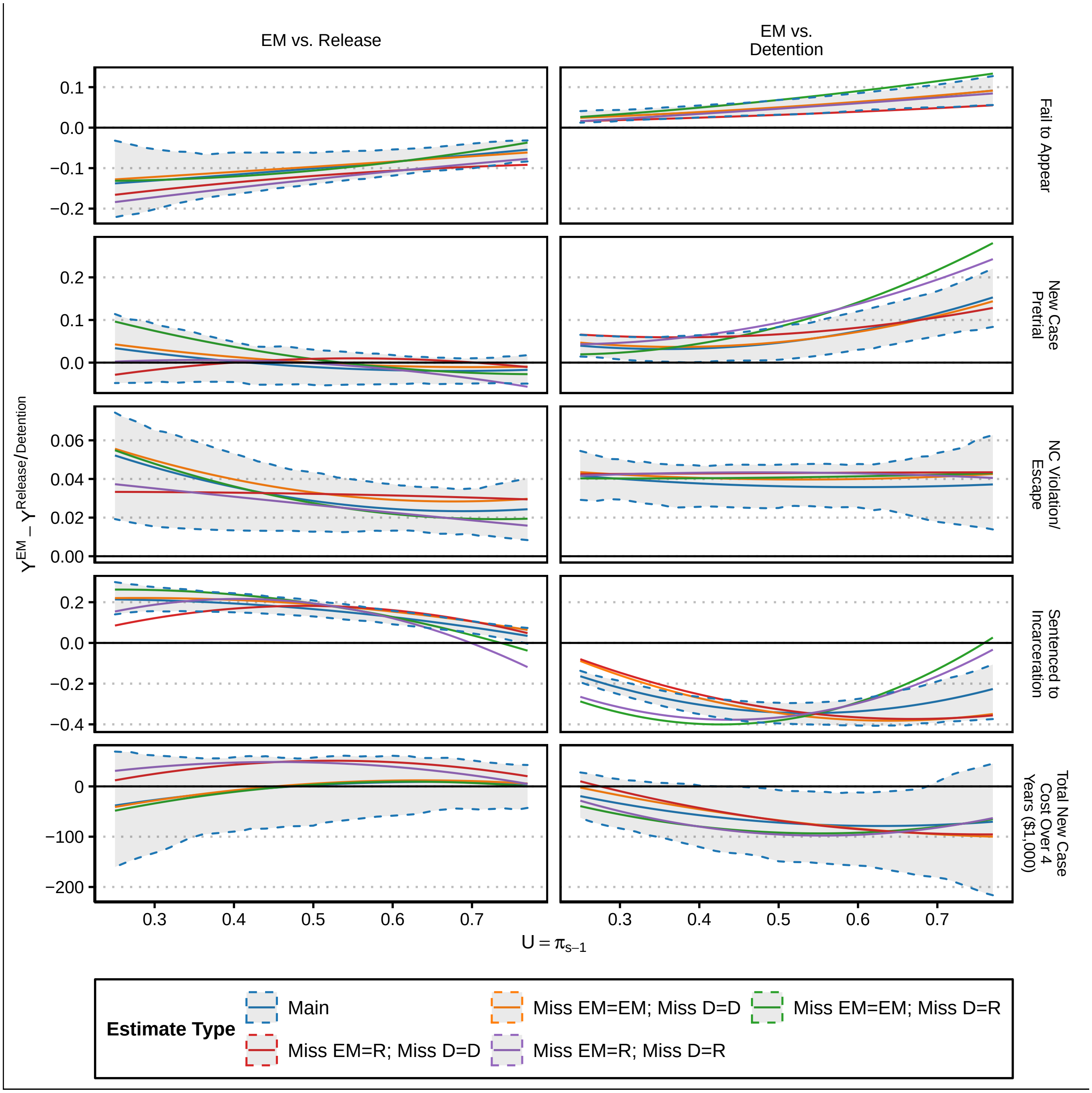


Figure A.23: MTEs for Recoded Missing Treatments

Note: Figure displays the marginal treatment effects (MTEs) for various robustness checks for the effect of EM relative to Release (left) and and EM relative to Detention (right) under different re-codings of pretrial treatments for cases with missing jail data. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of main estimates are computed using 200 bootstrap runs.

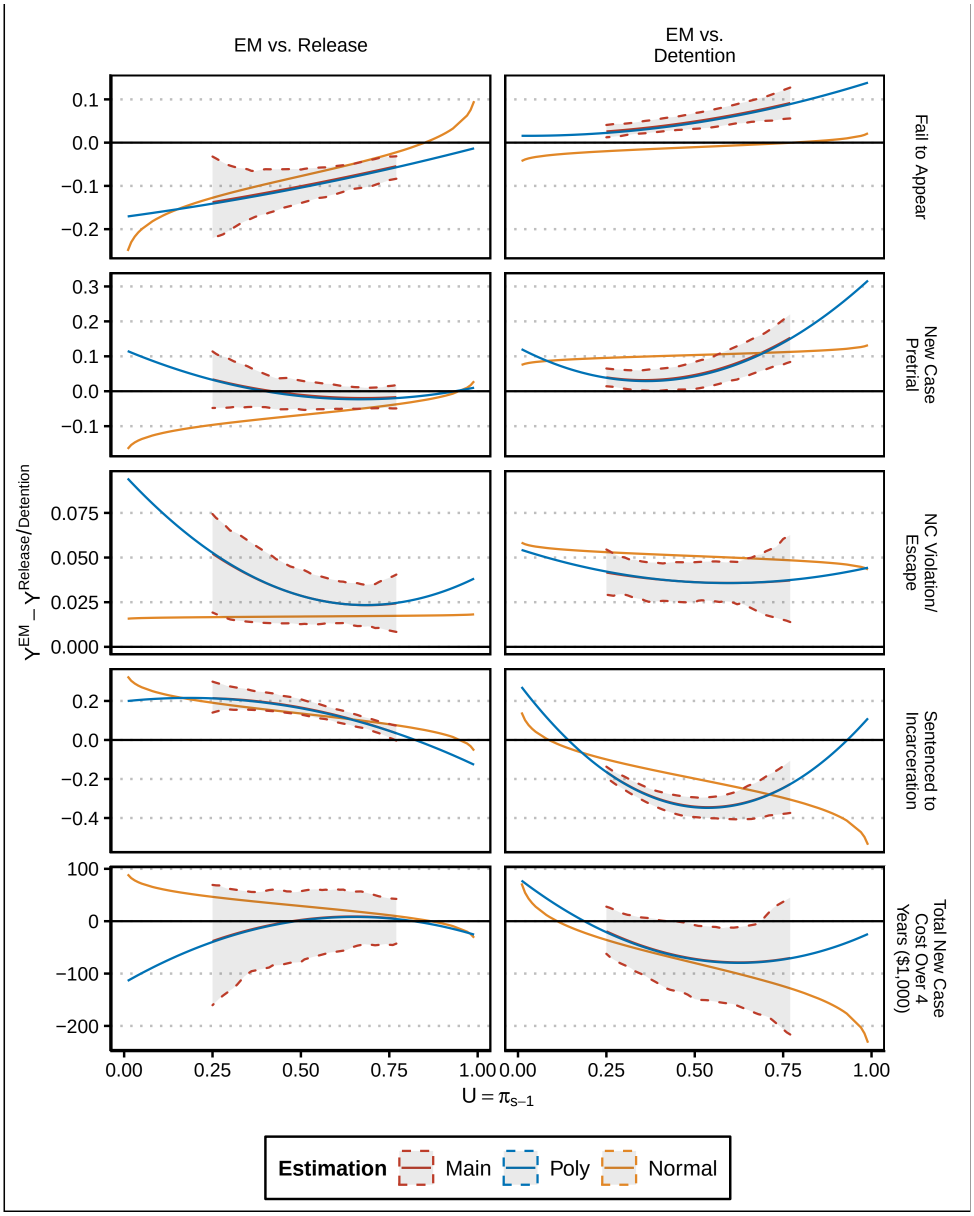


Figure A.24: MTEs for Comparing Main, Jointly Normal, and Polynomial Models

Note: Figure displays the marginal treatment effects (MTEs) across the distribution of defendant types (higher U means more unobservably resistant to higher treatment) assuming unobservables are drawn from a jointly normal distribution ('Normal'), following [104], the 'Main' specification over the common support (with confidence intervals), and a 3rd degree polynomials ('Poly') for Φ_s as a fully parametric model with full support.

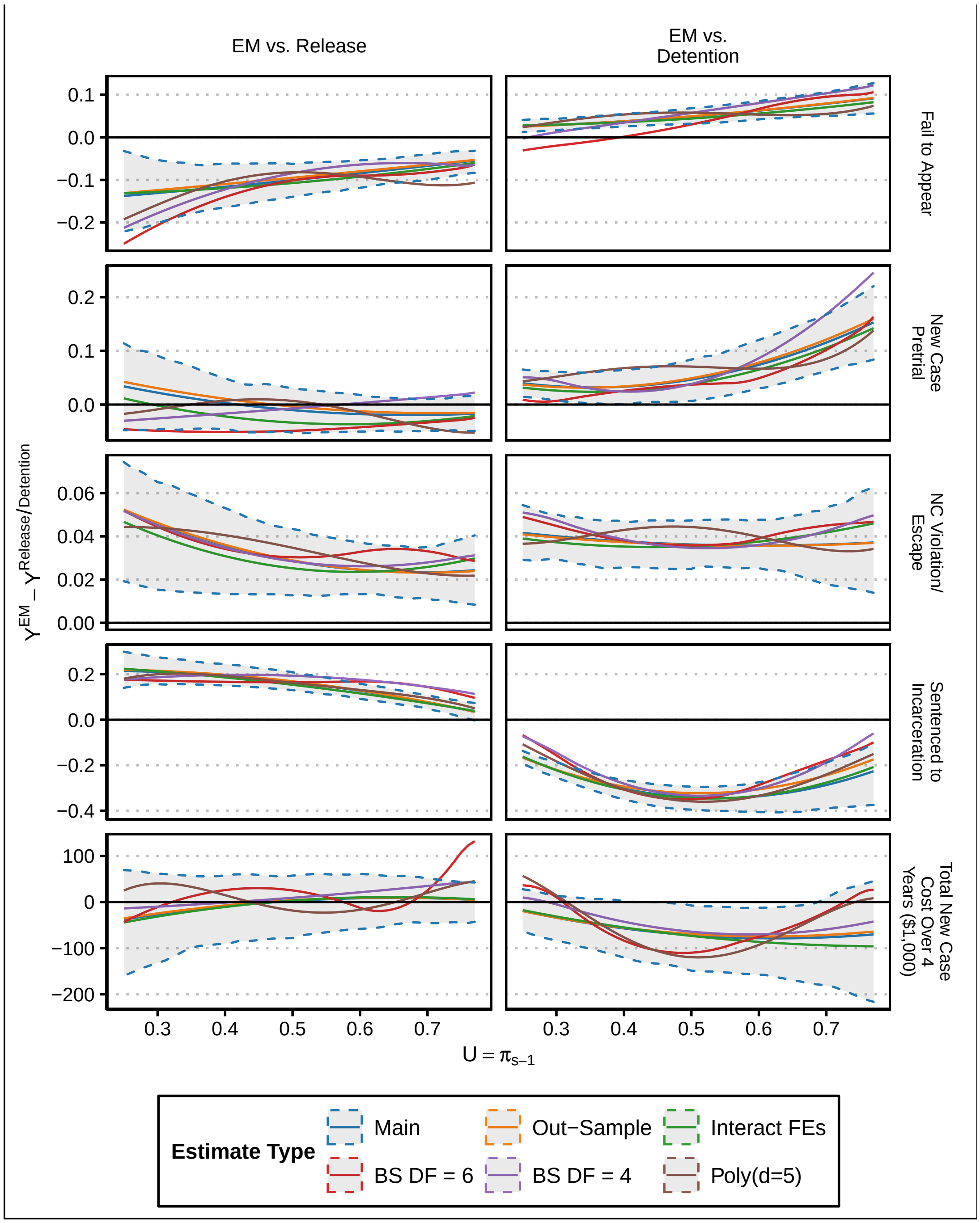


Figure A.25: MTEs with Alternative Specifications

Note: Figure displays the marginal treatment effects (MTEs) for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right) for different specifications. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of main estimates are computed using 200 bootstrap runs.

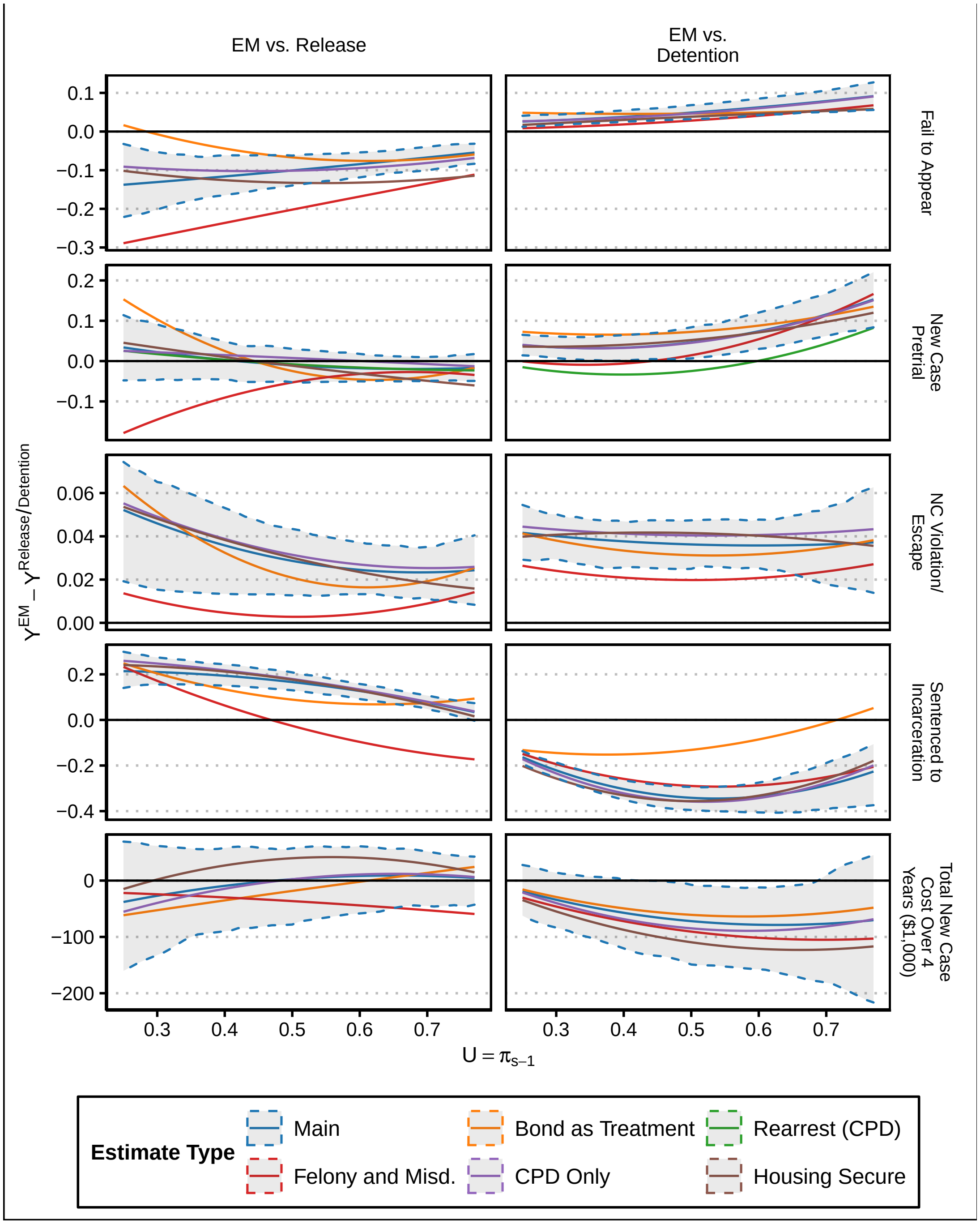


Figure A.26: MTEs with Alternative Samples

Note: Figure displays the marginal treatment effects (MTEs) for various robustness checks for the effect of EM relative to Release (left) and EM relative to Detention (right) for sub-samples of defendants. MTEs are recovered using the main semiparametric estimation method (equation (6) estimated with 3rd degree polynomial for Φ_s), unless otherwise specified. 95% confidence intervals of main estimates are computed using 200 bootstrap runs.

	$\hat{Z}^R$ (1 SD)	$\hat{Z}^D$ (1 SD)	F-Stat
	(1)	(2)	(3)
Pretrial Misconduct			
Fail to Appear	0.0042***	-0.0079***	51.88
	(0.00135)	(0.00111)	
New Case Pretrial	-0.0072***	-0.018***	51.069
	(0.00169)	(0.00179)	
New Case Pretrial - Serious	-0.0018*	-0.0062***	14.472
	(0.00106)	(0.00118)	
New Case Pretrial - Low-Level	-0.0036**	-0.0095***	18.212
	(0.00154)	(0.00158)	
New Case Pretrial - Violation/ Escape	-0.004***	-0.0081***	43.918
	(0.00072)	(0.00086)	
Case Outcomes			
Any Guilty Felony Charge	-0.0022	0.0018	2.502
	(0.00185)	(0.00152)	
Sentenced to Incarceration	0.0015	0.0054***	4.808
	(0.00165)	(0.00178)	
Post-Trial New Cases			
Total New Cases Post Trial within 4 Years	-0.014	-0.01	1.055
	(0.00999)	(0.01094)	
Total New Cases Post Trial within 4 Years - Felony	-0.0014	0.0018	0.216
	(0.00466)	(0.00486)	
Total Case Cost (Pre and Post) ($1,000)			
Total New Case Cost Over 4 Years	610	4400**	2.159
	(2701.76516)	(2174.83872)	
Total New Case Pretrial Cost	1600**	870	2.03
	(824.39783)	(725.40152)	
Total New Case Post-Trial Over 3 Years Cost	-1500	1900	1.141
	(2218.11026)	(1637.11968)	

Note: Table displays the results of reduced form regression, regressing the outcome on the two instruments, controlling for case observables and quarter and day of week fixed effects. Columns (1) and (2) contain the point estimate for the respective standardized instrument, and Column (3) contains the f-statistic from the projected model which already accounts for the influence of case observables and fixed effects. Standard errors clustered at the defendant level are in parantheses. ***p < 0.01; **p < 0.05; *p < 0.1

Table A.5: Reduced Form

	Release	Detain	Release	Detain
	(1)	(2)	(3)	(4)
Full Sample				
$Z^{Release}$	1***	0	1***	0
	(0.00954)	(0.00918)	(0.01938)	(0.01616)
Z^{Detain}	0	1***	0	1***
	(0.00862)	(0.00816)	(0.01595)	(0.01611)
Felony Sample				
$Z^{Release}$	0.73***	0.15***	1***	0.081***
	(0.0142)	(0.01408)	(0.02327)	(0.01897)
Z^{Detain}	-0.081***	1.1***	-0.00016	1.1***
	(0.00942)	(0.00983)	(0.01678)	(0.01856)
No Felony Sample				
$Z^{Release}$	1.1***	-0.16***	1.1***	-0.13***
	(0.0265)	(0.02308)	(0.03439)	(0.03002)
Z^{Detain}	0.013	0.73***	-0.047	0.86***
	(0.02839)	(0.02543)	(0.0395)	(0.03571)
Drug Charge Sample				
$Z^{Release}$	0.78***	0.14***	1***	0.1***
	(0.01994)	(0.01811)	(0.02773)	(0.02171)
Z^{Detain}	-0.028	1.1***	0.06***	1.1***
	(0.01784)	(0.01789)	(0.02259)	(0.02276)
Non-Black Sample				
$Z^{Release}$	0.96***	-0.07***	1***	-0.051*
	(0.02024)	(0.01724)	(0.03553)	(0.02881)
Z^{Detain}	0.0039	0.84***	0.026	0.85***
	(0.01871)	(0.01649)	(0.04073)	(0.03724)
Controls			X	X
Month + DoW Fes	X	X	X	X

Note: Table displays results of regressing the endogenous treatment (release or detain) on both instrumented treatment assignments with fixed effects and with controls in Columns (3) and (4) across predetermined subsamples. Standard errors clustered at the defendant level are in parantheses. ***p < 0.01; **p < 0.05; *p < 0.1

Table A.6: Average Monotonicity Test

A.3 General Results for MTEs

Marginal treatment effects (MTE) provide a structure for understanding how treatments affect individuals differently based on their unobservable characteristics. Generally, this involves estimating the effect of a treatment for defendants ranked according to their unobservable 'resistance' to treatment. In the binary case, this is straightforward, and there has been significant work on estimating MTEs for binary treatments.

When multiple ordered treatments are introduced, identification becomes more complicated theoretically and more difficult to estimate empirically. Prior work tends to make simplifying assumptions about the joint distribution of unobservables; following [102] (HUV) and [134] (HV), this generally a jointly normal distribution (as in [104]) which results in MTEs with pre-determined shapes (linear for most values of u). [101] is an exception, as they extend [291]'s bounding method to the case of ordered treatments by using discrete cutoffs as instruments to construct candidate marginal treatment response functions (MTRs) complying to specific shape restrictions, with the goal being bounding average treatment effects. In contrast, I focus on a general case, extending HUV, to point-identify TSMTEs and provide a method to estimate MTRs semiparametrically, relying on an interval of common support.

This section will focus on the case of estimating marginal treatment effects for ordered multi-valued treatments. The ordering of the treatments can be either cardinal — such as a dose-response function — or ordinal — where the treatments increase in intensity but have no clear quantitative distance. This section follows the work of HUV and HV closely. Building on their work, I relax one of HUV and HV's identification assumptions. The main contribution is to provide a tractable method for semiparametric estimation of MTRs.

A.3.1 Set up

Consider an individual (i, though individual subscripts are suppressed for brevity) either choosing between or being assigned to one of $\bar{S}$ different 'levels' of treatment which can be ordered by their intensity and given ranks such that $S \in \mathcal{S} = \{1, ..., \bar{S}\}$. For example, in the context of this paper, the individual is a defendant and the treatment levels are pretrial release ($S = 1$), EM ($S = 2$), and detention ($S = 3$), so $\bar{S} = 3$, but the levels could correspond to medicine dosage, years of schooling, or intensity of a social service intervention. Each individual has some set of observable (to the econometrician) characteristics X, as well as unobservable features $(\{\omega_1, ..., \omega_{\bar{S}}, V\})$. The unobservable factor V is observed by the agent who determines treatment.

Potential Outcomes For some outcome of interest, if the individual were assigned to treatment level $S = s$, their *potential* outcome is $Y_s = \mu_s(X, \omega_s)$. μ_s is some treatment-specific function, and X and ω_s is the individual-specific observable and unobservable components which contributes to Y_s. Let $D_s = 1$ if the defendant received treatment s, and $D_s = 0$ otherwise. From this, we know the observed outcome for an individual is:

$$Y = \sum_{s=1}^{\bar{S}} Y_s \times D_s = \sum_{s=1}^{\bar{S}} \mu_s(X, \omega_s) \times D_s.$$

Selection into Treatment The treatment level received by an individual is determined by a single index crossing a series of thresholds. The index, $T(Z, X)$, can be interpreted as the individuals 'net benefit' from a higher level of treatment (in the eyes of the agent assigning their treatment, possibly the individual themselves). The individual receives a treatment level higher than $S > s$ (e.g., $s + 1, s + 2, \ldots$) if and only if $T(Z, X) > C^s(W)$, where $C^s(W)$ is the cutoff value that is the highest level of $T(Z, X)$ which will result in being assigned to treatment level s. As before, X denotes observables that influence potential outcomes and possibly selection, while Z and W are observable instruments which do not directly influence outcomes but do influence selection either through the index (Z) or cutoffs (W).

From this, an individual is assigned to treatment level s if and only if:

$$D_s = 1(S = s) = 1[C^s(W) \geq T(Z, X) > C^{s-1}(W)]$$

where $C^{\bar{S}} = +\infty$ and $C^0 = -\infty$, because no one can receive a treatment higher than the highest level or lower than the lowest level (1), and $C^{s-1}(W) \leq C^s(W)\ \forall s$.

I assume that the single index can be decomposed into $T(Z,X) = \tau(Z,X) - V$ where $\tau(Z,X)$ is the individual's 'benefit' from a higher level of treatment and V is the individual's resistance to (or cost of) a higher level of treatment. The single index assumption has multiple implications. First, this separable form implies monotonicity, because for all individuals going from $\tau(Z,X)$ to $\tau(Z',X)$ will shift $T(\cdot,X)$ in the same way (similarly for if Z is fixed and X changes), and thus (weakly) move S in the same direction (HUV). Second, the existence of a single index, V, that determines treatment (conditional on observables) means that all factors can be reduced down to a single dimension in determining which treatment is optimal (to the agent deciding).

With this model, we seek to identify the effect of being assigned to treatment $s+1$ relative to treatment s for an individual with resistance level v and observables x. This is the transition-specific marginal treatment effect as coined in HV (TSMTE):

$$MTE_{s+1,s}(x,v) = \mathbb{E}[Y_{s+1} - Y_s | V = v, X = x]$$

A.3.2 Identification

In order to identify transition-specific marginal treatment effects, I assume the following assumptions from HUV (denoted HUV 1-6, though called OC 1-6 in the original paper):

Assumption 1. *HUV1:* $(\omega_s, V) \perp (Z, W)$ *for all* $s \in \mathcal{S}$ *conditional on* X.

Assumption 2. *HUV2:* $\tau(Z,X)$ *is a non-degenerate random variable conditional on* X *and* W.

Assumptions 1 and 2 ensure that the instruments are valid and relevant after conditioning on regressors.

Assumption 3. *HUV3: The distribution of* V *is absolutely continuous conditional on* X.

From Assumption 3, we can use probability integral transformation to get $U = F_V(V|X=x)$ which is uniformly distributed $U \sim Unif[0,1]$, conditional on X:

$$
\begin{aligned}
D_s &= 1[C^s(W) \geq T(Z,X) > C^{s-1}(W)] \\
&= 1[F_V(\tau(Z,X) - C^s(W)) \leq F_V(V) < F_V(\tau(Z,X) - C^{s-1}(W))] \\
&= 1[F_V(\tau(Z,X) - C^s(W)) \leq U < F_V(\tau(Z,X) - C^{s-1}(W))]
\end{aligned}
$$

Let $\pi_s(Z,X,W) = F_V(C^s(W) - \tau(Z,X)) = \Pr(S > s|Z,X,W)$. By construction, $\pi_0(Z,X,W) = 1$ and $\pi_{\bar{S}}(Z,X,W) = 0$. Then the selection equation becomes: $D_s = 1[\pi_s(Z,X,W) \leq U < \pi_{s-1}(Z,X,W)]$. With this, we can redefine the TSMTE with the selection unobservable being in terms of U (with a known distribution) rather than V (with an unknown distribution):

$$
MTE_{s+1,s}(x,u) = \mathbb{E}[Y_{s+1} - Y_s|U = u, X = x]
$$

Assumption 4. *HUV4:* $E|Y_s| < \infty \; \forall s \in S$

Assumption 5. *HUV5:* $0 < \Pr(S = s|X) < 1 \; \forall s \in S$

Assumption 6. *HUV6: The distribution of $C^s(W)$ conditional on X and Z and other $C^{s'}$ is non-degenerate and continuous* $\forall s \in \{1,...,\bar{S}-1\}$.

With these assumptions, HUV and HV show that the TSMTE is identified by taking: $\frac{\partial \mathbb{E}[Y|\pi(Z,X,W)=\pi,X=x]}{\partial \pi_s} = MTE_{s+1,s}(u,x) = \mathbb{E}[Y_{s+1} - Y_s|U = u, X = x]$, where $\pi = [\pi_1,...,\pi_{\bar{S}-1}]$.

However, this introduces complications that make semiparametric or nonparametric estimation of such a model difficult, particularly with more than 3 treatments, for two main reasons. First, Assumption 6 effectively requires variation in the C^s's conditional on all other $C^{s'} \; \forall s' \in \{1,...,\bar{S}-1\} \setminus s$ and observables wherever the TSMTE is to be identified. This means if there are $\bar{S} = 6$ treatment levels, then wherever one wishes to estimate $MTE_{4,3}(u)$, there must be variation in all 5 π_s's. This may very well not be the case, for example, at high u comparing $S = 4$ to $S = 3$, π_1 or π_2 may be degenerate (or effectively so in the data). In this scenario, Assumption 6 would not hold.

Second, the form $\frac{\partial \mathbb{E}[Y|\pi(Z,X,W)=\pi]}{\partial \pi_s}$ conditions on vector $\pi = [\pi_1,...,\pi_{\bar{S}-1}]$ for estimation and taking a partial derivative of the function $\mathbb{E}[Y|\pi]$ with respect to the specific π_s of interest, and it does not allow us to recover treatment responses only treatment effects. Assuming both ω_s and U are drawn from a jointly normal distribution is a common method for estimation, though this fully parametric assumption leads to effectively linear TSMTEs for most values of u.

I provide an alternative identification method and a weaker assumption to replace Assumption 6, which improves upon both of these limitations. First, rather than recovering TSMTEs through a local-IV approach, we can recover TSMTEs as the difference between marginal treatment response (MTR) functions ([94], [95], [291], [101]) at a fixed value of $X = x$ and $U = u$:[1]

$$MTE_{s+1,s}(u, x) = \mathbb{E}[Y_{s+1}|U = u, X = x] - \mathbb{E}[Y_s|U = u, X = x]$$

Identification of an MTR (e.g., $\mathbb{E}[Y_s|U, X]$) can be achieved relying solely on variation in adjacent π_s's (e.g., π_s and π_{s-1}). Specifically, I provide an weaker alternative to Assumption 6, which both reduces the required variation for identification of TSMTEs when $\bar{S} > 3$ and allows the identification of MTRs and simpler estimation using the separate approach:

Assumption 7. *For all $s \in \mathcal{S} \setminus \{1, \bar{S}\}$, the joint distribution of $\pi_s(Z, X, W)$ and $\pi_{s-1}(Z, X, W)$ conditional on X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. Furthermore, the joint distribution of $\pi_s(Z, X, W)$ and $\pi_{s-1}(Z, X, W)$ conditional on X is non-degenerate in the sense that its support cannot be reduced to a subset on $\mathbb{R}$.*

Assumption 7 improves upon Assumption 6. First, because it only makes assumptions on the joint distribution of π_s and π_{s-1}, simply requiring that that they are not highly codependent conditional on X — in the language of Assumption 6, for $MTE_{s+1,s}$, C^s only need be non-degenerate and continuous conditional on C^{s+1}, C^{s-1}, and X. Second, this assumption lends itself to a simple semiparametric estimation approach, as will be discussed below, and thus is more feasible for applications in applied work. TSMTEs are then the difference between MTRs, and MTRs are identified under Assumptions (1)-(5) and (7):

[1] The framework stems from the literature on estimating the marginal distributions of potential outcomes in [292] and [293].

Theorem 1. *Let Assumptions (1)-(5), and (7) hold. Then,*

$$\mathbb{E}[Y_1|U=u, X=x] = -\frac{\partial \mathbb{E}[Y \times D_1|\pi_1(Z,X,W)=\pi_1, X=x]}{\partial \pi_1}\bigg|_{\pi_1=u}$$
$$\mathbb{E}[Y_{\bar{S}}|U=u, X=x] = \frac{\partial \mathbb{E}[Y \times D_{\bar{S}}|\pi_{\bar{S}-1}(Z,X,W)=\pi_{\bar{S}-1}, X=x]}{\partial \pi_{\bar{S}-1}}\bigg|_{\pi_{\bar{S}-1}=u}$$

And, for all $s = 2, \ldots, \bar{S}-1$,

$$\mathbb{E}[Y_s|U=u, X=x] = \frac{\partial \mathbb{E}[Y \times D_s|\pi_{s-1}(Z,X,W)=\pi_{s-1}, \pi_s(Z,X,W)=\pi_s, X=x]}{\partial \pi_{s-1}}\bigg|_{\pi_{s-1}=u}$$
$$= -\frac{\partial \mathbb{E}[Y \times D_s|\pi_{s-1}(Z,X,W)=\pi_{s-1}, \pi_s(Z,X,W)=\pi_s, X=x]}{\partial \pi_s}\bigg|_{\pi_s=u}.$$

Proof of Theorem 1. Using Assumptions (1)-(5) and (7), write (suppressing W):

$$\begin{aligned}
&\mathbb{E}[Y \times D_s|Z=z, X=x, W=w]\\
&= \mathbb{E}[\mu_s(X,\omega_s)1\{\pi_{s-1}(Z,X,W) > U \geq \pi_s(Z,X,W)\}|Z=z, X=x, W=w]\\
&= \mathbb{E}[\mu_s(X,\omega_s)1\{\pi_{s-1}(Z,X,W) > U \geq \pi_s(Z,X,W)\}|X=x] \textit{ by (1)}\\
&= \int_{\pi_s(z,x,w)}^{\pi_{s-1}(z,x,w)} \mathbb{E}[Y_s|U=u, X=x]du \textit{ by (3)}
\end{aligned}$$

Then, because

$$\mathbb{E}[Y \times D_s|Z=z, X=x, W=w] = \int_{\pi_s(z,x,w)}^{\pi_{s-1}(z,x,w)} \mathbb{E}[Y_s|U=u, X=x]du,$$

taking the derivative of both sides with respect to π_{s-1} (when $s > 1$) gives:

$$\left.\frac{\partial \mathbb{E}[Y \times D_s | Z = z, X = x, W = w]}{\partial \pi_{s-1}}\right|_{\pi_{s-1}=u} = \mathbb{E}[Y_s | U = u, X = x],$$

and similarly taking the derivative of both sides with respect to π_s gives:

$$\left.\frac{\partial \mathbb{E}[Y 1(S = s) | Z = z, X = x, W = w]}{\partial \pi_s}\right|_{\pi_s=u} = -\mathbb{E}[Y_s | U = u, X = x].$$

Because $D_s = 1(S = s) = 1[\pi_{s-1}(Z, X, W) > U \geq \pi_s(Z, X, W)]$ and by Assumption 7, only variation in the adjacent π's are relevant. So:

$$\begin{aligned} &\left.\frac{\partial \mathbb{E}[Y \times D_s | Z = z, X = x, W = w]}{\partial \pi_{s-1}}\right|_{\pi_{s-1}=u} \\ = &\left.\frac{\partial \mathbb{E}[Y \times D_s | \pi_{s-1}(Z, X, W) = \pi_{s-1}, \pi_s(Z, X, W) = \pi_s, X = x]}{\partial \pi_{s-1}}\right|_{\pi_{s-1}=u} \end{aligned}$$

and

$$\begin{aligned} &\left.\frac{\partial \mathbb{E}[Y \times D_s | Z = z, X = x, W = w]}{\partial \pi_s}\right|_{\pi_s=u} \\ = &\left.\frac{\partial \mathbb{E}[Y \times D_s | \pi_{s-1}(Z, X, W) = \pi_{s-1}, \pi_s(Z, X, W) = \pi_s, X = x]}{\partial \pi_s}\right|_{\pi_s=u} \end{aligned}$$

Finally, we have for s with non-degenerate π_s and π_{s-1}:

$$\begin{aligned} \mathbb{E}[Y_s | U = u, X = x] &= \left.\frac{\partial \mathbb{E}[Y \times D_s | \pi_{s-1}(Z, X, W) = \pi_{s-1}, \pi_s(Z, X, W) = \pi_s, X = x]}{\partial \pi_{s-1}}\right|_{\pi_{s-1}=u} \\ &= -\left.\frac{\partial \mathbb{E}[Y \times D_s | \pi_{s-1}(Z, X, W) = \pi_{s-1}, \pi_s(Z, X, W) = \pi_s, X = x]}{\partial \pi_s}\right|_{\pi_s=u}. \end{aligned}$$

□

Assumptions 2, 4, and 5 ensure values in the population are well-defined, while Assumption 7 ensures variation in adjacent π_s's.

With this, we can identify the TSMTE between levels s and $s+1$ by identifying the conditional means (MTRs) for s and $s+1$:

$$MTE_{s+1,s}(x,u) = \mathbb{E}[Y_{s+1} - Y_s|U = u, X = x] = \mathbb{E}[Y_{s+1}|U = u, X = x] - \mathbb{E}[Y_s|U = u, X = x]$$

with

$$\begin{aligned} &\mathbb{E}[Y_s|U = u, X = x] \\ &= \left.\frac{\partial \mathbb{E}[Y_s \times D_s|\pi_{s-1}(Z,X,W) = \pi_{s-1}, \pi_s(Z,X,W) = \pi_s, X = x]}{\partial \pi_{s-1}}\right|_{\pi_{s-1}=u} \\ &= -\left.\frac{\partial \mathbb{E}[Y_s \times D_s|\pi_{s-1}(Z,X,W) = \pi_{s-1}, \pi_s(Z,X,W) = \pi_s, X = x]}{\partial \pi_s}\right|_{\pi_s=u}. \end{aligned} \tag{A.1}$$

And this equality applies only to $1 < S < \bar{S}$ — so for intermediate treatment levels, $\mathbb{E}[Y_s|U = u, X = x]$ is over-identified.

While $\mathbb{E}[Y_s \times D_s|\pi_{s-1}(Z,X,W) = \pi_{s-1}, \pi_s(Z,X,W) = \pi_s, X = x]$ can be estimated non-parametrically, in practice the data requirements make such estimation are rarely feasible, and semiparametric estimation is often the preferred approach in practice in the MTE literature.

For semiparametric estimation, I assume that $\mu_s(X,\omega_s)$ is composed of additively separable functions of X and ω_s, essentially that across all values of covariates, the effect of unobservables works the same and it allows for treatment effects on observables and unobservables separately ([100]). This assumption (either directly or as a result of full independence) is common in the literature ([133], [86], [95], [294], [101]). Specifically, I assume: $Y_s = \beta_s X + \omega_s$ and $\mathbb{E}[Y_s|x,u] = \beta_s x + \mathbb{E}[\omega_s|U = u]$. Then, the marginal treatment effect of moving from one treatment to the next highest one (s to $s+1$) is:

$$\begin{aligned} MTE_{s+1,s}(U = u, X = x) &= \mathbb{E}[Y_{s+1} - Y_s|U = u, X = x] \\ &= (\beta_{s+1} - \beta_s)x + \mathbb{E}[\omega_{s+1}|U = u] - \mathbb{E}[\omega_s|U = u] \end{aligned}$$

A.3.3 Semi-Parametric Estimation of MTRs

The following section will provide simple functional form assumptions and an accompanying semiparametric estimation procedure. For notational purposes, I suppress W and allow it to be subsumed by Z, X, as in [104]. In practice, when estimating π_s's there is no explicit distinction between index instruments (Z) and cutoff instruments (W). Assumptions for estimation are stronger than those for identification above (see Appendix A.3.3).

Estimation Form

With additive separability, we can recover $\mathbb{E}[Y_s|X = x, U = u]$, which we do not observe, by starting with $Y \times D_s$, which we do observe. Specifically:

$$\begin{aligned} &\mathbb{E}[Y \times D_s|\pi_{s-1}(Z, X) = \pi_{s-1}, \pi_s(Z, X) = \pi_s, X = x] \\ &= \beta_s x(\pi_{s-1} - \pi_s) + \Lambda_s(\pi_{s-1}) - \Lambda_s(\pi_s) \end{aligned} \tag{A.2}$$

where each $\Lambda_s(k) = \int_0^k \mathbb{E}[\omega_s|U = u]du$. In practice, we can approximate each $\Lambda_s(k)$ with sieves, such that $\Phi_s(k)$ is a vector of basis functions and ϕ_s is a vector of coefficients:

$$\Lambda_s(k) - \Lambda_s(k') \approx \phi_s'[\Phi_s(k) - \Phi_s(k')] = \sum_{j=1}^{J} \phi_{s,j}(k)[\vartheta_{s,j}(k) - \vartheta_{s,j}(k')].$$

Assume also that we have i.i.d. data $\{(Y_i, S_i, Z_i, X_i) : i = 1, \ldots, n\}$.

Then, from equation (A.1) and using the functional form assumption in equation (A.2), for all s such that $1 < s < \bar{S}$:

$$\mathbb{E}[Y_s|X = x, U = \pi_s \text{ or } \pi_{s-1}] = -(-\beta_s x - \frac{\partial}{\partial \pi_s}\Lambda_s(\pi_s)) = \beta_s X + \frac{\partial}{\partial \pi_{s-1}}\Lambda_s(\pi_{s-1})$$

and for $s = 1$:

$$\mathbb{E}[Y_1|X = x, U = \pi_1] = \beta_1 x + \frac{\partial}{\partial \pi_1}\Lambda_1(\pi_1)$$

while for $s = \bar{S}$:

$$\mathbb{E}[Y_{\bar{S}}|X = x, U = \pi_{\bar{S}-1}] = \beta_{\bar{S}} x + \frac{\partial}{\partial \pi_{\bar{S}-1}}\Lambda_{\bar{S}}(\pi_{\bar{S}-1})$$

Estimation Steps

Estimation is based on equation (A.2). We can approximate Λ_s using B-splines or a polynomial of π_s (and similarly for π_{s-1}). The main estimation procedure in this paper follows five steps:

- 1. Recover estimates of π_s ($\forall s \in \{1, ..., \bar{S}-1\}$) as probabilities (i.e., $\hat{\pi}_s \in [0,1]$) (for example, using separate probit or logistic regressions) by regressing the treatment level being higher than s on instruments and regressors:

$$1\{S_i > s\} = \beta^s [X_i, Z_i, W_i] + \epsilon_i^s$$

  Then predict $\hat{\pi}_s$ $\forall s \in \{1, ..., \bar{S}-1\}$. In this paper, I use a probit specification regressing $s \in [EM, Detention]$ ($s > 1$) and $s = Detention$ ($s > 2$) on judge fixed effects interacted with observables and time fixed effects to get predicted values for $\hat{\pi}_1$ and $\hat{\pi}_2$, respectively.

- 2. For each $s \in S$, construct $\Phi_s(\pi_s)$ and $\Phi_s(\pi_{s-1})$ either as polynomials or B-splines, each being a vector of basis functions.

- 3. For each $s \in S$, regress

$$Y \times D_s = \beta_s X(\pi_{s-1} - \pi_s) + \phi_s(\Phi_s(\pi_{s-1}) - \Phi_s(\pi_s)).$$

  If $s = 1$ then exclude $\Lambda_s(\pi_{s-1})$, and if $s = \bar{S}$ then exclude $\Phi_s(\pi_s)$. ϕ_s is a vector of coefficients with each element corresponding to each basis function. For example, if we are using a 3rd degree polynomial, then $\Phi_s(k) = [k, k^2, k^3]$, so $\phi_s = [\phi_s^1, \phi_s^2, \phi_s^3]$ and $\Phi_s(\pi_{s-1}) - \Phi_s(\pi_s) = [\pi_{s-1} - \pi_s, \pi_{s-1}^2 - \pi_s^2, \pi_{s-1}^3 - \pi_s^3]'$.

- 4. Then compute the estimate for $\mathbb{E}[Y_s|x, u]$ as:

$$\widehat{\mathbb{E}[Y_s|x,u]} = \begin{cases} \textit{if } s = 1 & \hat{\beta}_s x + \hat{\phi}_s \Phi_s'(\pi_s),\ u = \pi_s \\ \textit{if } \ s > 1 & \hat{\beta}_s x + \hat{\phi}_s \Phi_s'(\pi_{s-1}),\ u = \pi_{s-1} \end{cases}$$

In the $s = 1$ case, $\widehat{\mathbb{E}[Y_s|x,u]} = -\frac{\partial \widehat{\mathbb{E}[Y(S=s)|x,u=\pi_s]}}{\partial \pi_s} = -[-\beta_s x - \hat{\phi}_s \Phi'_s(\pi_s)] = \beta_s x + \hat{\phi}_s \Phi'_s(\pi_s)$.

- 5. For each value of u in the support of both s and $s+1$ and any value of $X = x$, compute $\widehat{MTE}_{s+1,s}(x,u) = \widehat{\mathbb{E}[Y_{s+1}|x,u]} - \widehat{\mathbb{E}[Y_s|x,u]}$.

Assumptions for Estimation

Assumption 8. *E1:* $(\omega_s, U) \perp (Z, X)$ *for all* $s \in \mathcal{S}$.

Assumption 9. *E3: The distribution of U is uniform on* $[0, 1]$.

Assumption 10. *E4:* $E|Y_s| < \infty \; \forall s \in S$

Assumption 11. *E5:* $0 < \Pr(S = s) < 1 \; \forall s \in S$

Assumption 12. *E6: For all* $s \in \mathcal{S} \setminus \{1, \bar{S}\}$*, the joint distribution of* $\pi_s(Z, X)$ *and* $\pi_{s-1}(Z, X)$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^2$*. Furthermore, the joint distribution of* $\pi_s(Z, X)$ *and* $\pi_{s-1}(Z, X)$ *is non-degenerate in the sense that its support cannot be reduced to a subset on* $\mathbb{R}$*.*

Assumption 8 replaces V with U and W is subsumed by X, Z and imposes a full independence assumption (rather than conditional independence). From this, the other assumptions no longer condition on X (due to full independence), W is subsumed into X, Z, and U is used in place of V for expediency removing the need for assumptions on τ.

A.3.4 Common Support

Estimation requires variation in $\pi_{s-1} - \pi_s$, which provides variation in higher terms of $\Lambda_{s-1} - \Lambda_s$. This must be factored into determining where the MTRs are estimable, in addition to ensuring that the MTRs of two adjacent treatments exist. Common support also determines the region over which we are able to produce counterfactuals and treatment effects.

Given that $\pi_0 = 1$ and $\pi_3 = 0$ for all individuals, the variation of π_1 and π_2 are of most concern — though their correlation is high (0.67) this is to be expected given that if $s > 2$ then $s > 1$ as well. As shown in Figure A.7, below the 45 degree line (such that $\pi_2 < \pi_1$), there is variation in values of each π excluding very high values of π_1 and very low values of π_2. The right panel in Figure A.7 displays the same plot for the felony-only sample.

To estimate the MTEs however, we require common support for the π's and variation within a treatment level for which we will estimate a marginal treatment response function. Figure A.8 displays the distribution of $\pi_{s-1} - \pi_s$ for each treatment level (right panel displays the same for felony-only sample). For $s = 1, 3$, $\pi_{s-1} - \pi_s$ covers the entire unit interval, but for $s = 2$, it falls short, with strong support only between about $[0.05, 0.8]$.

Figure A.9 displays the supports for each relevant π by treatment type (i.e., π_1 for $S = 1, 2$, and π_2 for $S = 2, 3$) for the main sample, along with 1% trim vertical lines denoting the 1st and 99th percentiles within the treatment-specific sample. There is nearly full support for $S = 1, 3$. However, for $S = 2$, the support ranges from about $\pi_1 \in [0.25, 1]$ and $\pi_2 \in [0.1, 0.77]$. Since the overlap we require is where there is common support for π_1 and π_2 for $S = 2$, we can only compute the MTR for EM for $\pi_1, \pi_2 \in [0.25, 0.77]$, which limits the range we can compute TSMTEs for EM versus release and EM versus detention, as EM is the most limited in support.

A.3.5 Confidence Intervals

The main results use bootstrapped MTRs based on 200 runs to compute confidence intervals. For each run, data is sampled with replacement and MTRs are computed. π's are not recalculated each run. Then MTEs are the difference between MTRs. Each run provides MTE estimates for each value of π_1 and π_2 for EM versus release and EM versus detention, respectively, and the 95% confidence intervals are taken as the 97.5% and 2.5% (195th and 5th highest values) for each MTE point independently — meaning if bootstrap sample 1 corresponds to the 2.5% value for $MTE_{EM,R}(\pi_1 = 0.5)$, it does not mean the 2.5% value for $MTE_{EM,R}(\pi_1 = 0.51)$ is from bootstrap sample 1 as well. As a result, the confidence intervals are not symmetric, as would result if standard errors were computed using the distribution of estimates.

While the confidence intervals in the main results do not account for the fact that π_1 and π_2 are estimated objects, doing so has very little effect on the size of confidence intervals. I test this by using 400 bootstrap runs in which π_1 and π_2 are re-computed each run using the full (bootstrapped) sample then filtering for main felony sample observations. This full bootstrap only increase the 95% confidence interval sizes for the five main robustness outcomes on the parameters of interest (CATEs, CATR, and CATD) by an average of 1.04% with a maximum increase of 18%.

Appendix B: Chapter 2

B.1 Additional Background and Results

B.1.1 Entrance into the CPD Police Academy

In order to become an officer in the CPD, applicants must first meet multiple qualifications before applying to take the entrance exam. For example, by the time of starting at the academy, one must be a US citizen, a resident of Chicago, have sufficient credit hours at a college or university, and meet the age requirement ([191]). Potential applicants meeting these qualifications can apply to take the CPD entrance exam, and they will be notified of the test date and location after the application period ends ([188]).[1]

Applicants who pass the written exam are then assigned a random lottery number indicating the order in which they will be called into the academy. Random assignment to the academy was not always the case; it was introduced in the early 1990's in an attempt to increase diversity ([296]). After an applicant's number is drawn, they must pass a background check, drug screening, and medical, psychological and physical exams ([191]). Upon passing these requirements, potential officers are admitted into the academy.

There are usually tests once every 2 or 3 years (not including makeup exams)—but in 2006 there were four exams issued.[2] Generally, thousands of people take the CPD's written exam and a large portion of them meet the minimum passing score (see Figure B.2.1). Given the large number of passing applicants, many do not ever have their numbers called before the applicant list is retired. Despite my best efforts, I have not been able to obtain any indication of when the applicant lists

[1] As late as the 2013 exam, veterans began to received preference in their lottery numbers– though this is not well defined in documentation. However, this preference is unlikely to be important considering almost all (over 95%) of recruits have military experience in the full sample. This very large amount of veterans is consistent with more recent estimates from the Office of the Inspector General ([295]).

[2] One is labeled a '2005' exam in Figure B.2.1, but it took place in February 2006.

are retired (according to the CPD such documentation may not even exist). Also, applicants from a test are likely to be admitted possibly years after they took the test initially, and their entrance into the academy likely occurs while more applicants are taking a new test. This makes identifying which cohorts come from which tests (i.e. the pool from which officers are randomly assigned) difficult.

To the best of my knowledge, the main sample (July 2012 to May 2014) cohorts are an exception, and these cohorts all came from the same exam issued in December of 2010 (see Figure B.2.1). The December 2010 exam was the last exam issued before the December 2013 exam. The only sizable cohort to enter in 2011 was on October 17, 2011, then about 8 months pass until the first sizable cohort of 2012 started on July 02, 2012. Following this, there were a total of 7 sizable cohorts starting between July and December of 2012. Then, there is continuous intake of cohorts until May of 2014, when there is a 3 month gap until the next cohort. Given that it takes time for the CPD to draw in passing recruits and give them their multiple examinations, I believe that the main sample cohorts were all drawn from the December 2010 exam.

Further supporting this is the change in the composition of cohorts before and after 2012. As shown in Panel A of Figure B.1, the 2011 cohort has a higher share Black than almost every cohort in the 2012-2014 period, while it is within the range of the Exam 2006 cohorts (likely drawn from the 2006 tests). Similar patterns emerge when looking at share of the cohort which speaks Spanish (see Panel B of Figure B.1), where all of the 2006 cohorts have strictly smaller shares of Spanish speakers compared with any 2010 cohort. Finally, minimum start age (Panel C) increases successively for each of the pre-2010 cohorts (as expected since these recruits have been waiting at least 4 years to enter), while it decreases slightly in the first 2010 cohort and significantly in the second 2010 cohort. Anecdotally, an officer I spoke with who started the academy in 2012 confirmed that their cohort was comprised of 2010 test takers.

In separating the Exam 2006 cohorts (starting in 2009 and ending in 2011) from the Exam 2010 cohorts, and determining if all Exam 2006 cohorts actually came from the four 2006 exams (and not the 2004 exam), I use posts on a police forum (`https://forum.officer.com/`)

in 2009, 2010, and 2011. One poster on November 17th, 2009, states: "Just got the call… the academy starts December 16th… My number is 1036, and I am a June 06 tester." ([297], pg.29). December 16th, 2009, is the start date of the first cohort in my full sample. This is followed by a flurry of other posters stating their numbers also got called for the same start date. The only cohort before it was in March of 2009, which according to a poster in on March 6th 2009, "From what I know [the March 2006 cohort] it's a mix of Feb 06 and early June 06 testers." ([297], pg.9). Overall, this indicates the 2009 and 2010 cohorts came from Exam 2006 test takers only.

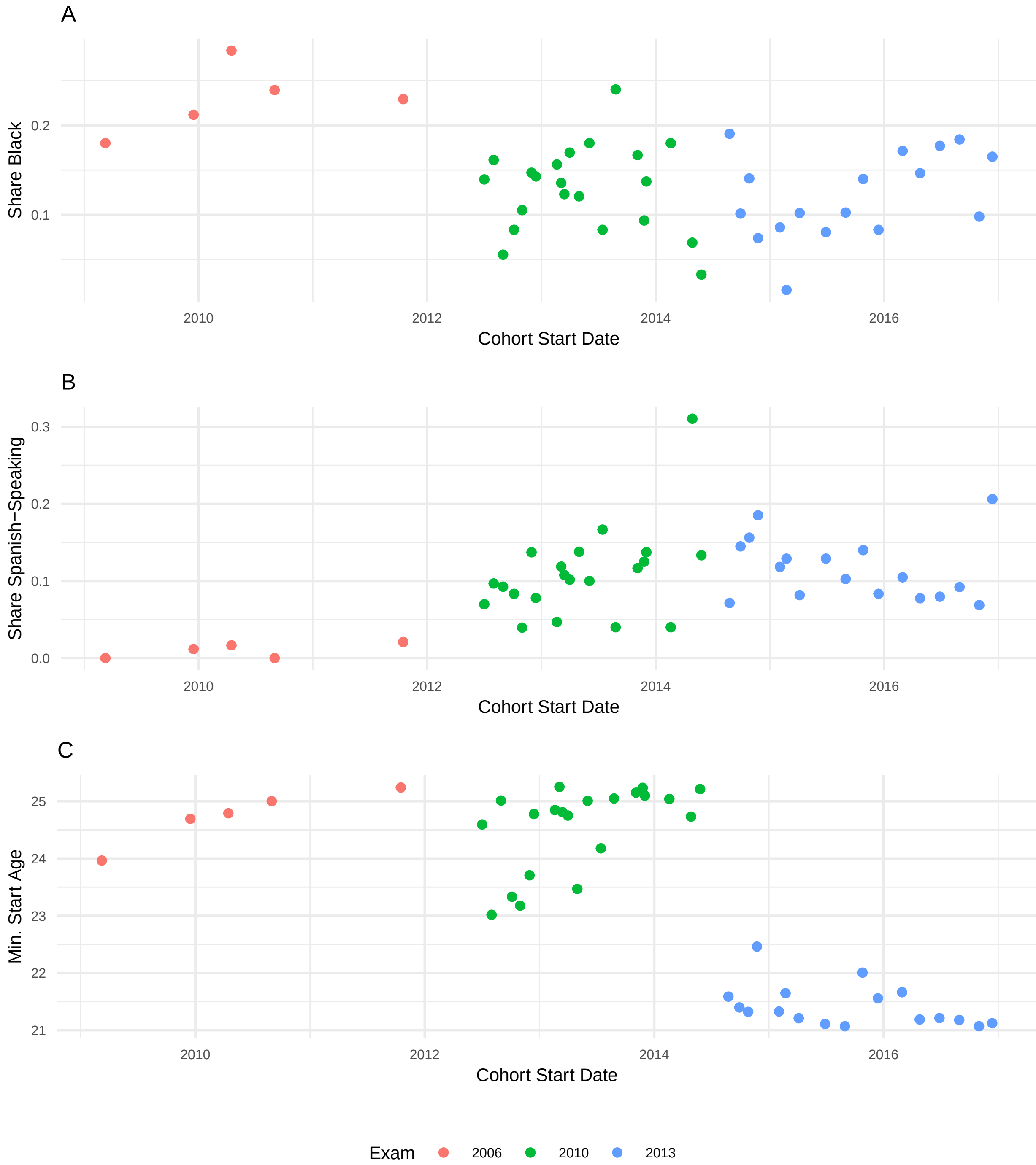
Next, the main question is did the single 2011 (in October) cohort end the Exam 2006 cohorts or start the Exam 2010 cohorts? According to a different thread on the same site, a poster on December 4th, 2010, states: "With roughly 40 candidates ready for hire off the 2006 test, and a new test next week, its about time we started this thread. For those who are wondering, the last of the 2006 list (40 people) were scheduled to start on 01 November [2010] but according to my BI who I call twice every month, the class has been pushed back and only the fine folks at city hall know the date. In my humble opinion city hall is waiting on the new year [2011] to start our class because of the new budget and the new pension system for new hires." ([298], pg.1) On September 30th, 2011, a poster states that their cohort ("2011-1") will "soon fill the halls of the Chicago Police Academy" ([298], pg. 6), and another poster, on October 18th, 2011, (one day after the 2011 cohort starts in the data) states that the class has "About 50" recruits (49 in the data). The rest of this forum discusses the composition of this cohort. It is stated that this cohort will exhaust the rest of the 2006 applicants (at least 32) and fill the rest either with 2006 applicants who won appeals or 2010 testers. So, based on these discussions, the single 2011 cohort finished off the 2006 Exam cohorts, and was potentially mixed with a small number of 2010 Exam takers– though this seems to be an unusual practice and only a result of the small number of potential recruits in the 2006 tests ([298], pg. 3). Because of this issue with mixing a single cohort, Table B.2.6 displays the results of Table 2.4 excluding the 2011 cohort, and the results are highly similar.

After May 2014, the cohorts until December 2016 (the last cohort I use in the extended sample) are from the 2013 test. The 2013 test recruits had the new feature that they were permitted to begin the academy at the age of 21, lower than the previous requirement of 23 ([191]). As can be seen in Panel C of Figure B.1, the lowest starting age per cohort drops to 21 after the May 2014 cohort. Thus, I can distinguish between the 2010 and 2013 test cohorts using this feature. The end of the 2013 test cohorts occurs after the final cohort in the full sample in December of 2016. Even though there was a test issued in April 2016, based on forum posts about 2016 recruitment the 2016 test-takers had not begun to be drawn in by the end of 2016. Following many 2016 test takers wondering when their cohorts would be drawn in, one poster stated on December 26, 2016, "People that took the exam in 2013 are still being processed. I believe about 9k people passed the written exam this year" ([299], pg. 138). So, I am confident that the Exam 2013 sample does not contain 2016 test cohorts. Based on the panels in Figure B.1, there is fairly consistent cohort composition across the Exam 2013 cohorts. While extending my cohorts beyond December 2016 is possible, because my panel data extends to 2018 (overlapping with court data and outcomes), including the first cohorts in 2017 would not contribute much to my analysis as these officers would have less than 6 months of observations in the panel data after their probationary period.

As I am less confident as I move away from the Exam 2010 cohorts – not knowing the beginning of the Exam 2006 cohorts and being restricted in the panel with post-2016 cohorts – I focus on them (the main sample) in my analysis. However, as incorporating the Exam 2006 and Exam 2013 cohorts into my analysis provides results which are generally consistent with my main sample results and they provide a significant increase in sample size, I use them collectively as well.

B.1.2 Random Assignment

Given that the timing of when a recruit can enter the academy is determined by a random lottery number, cohorts are as-good-as-randomly assigned, and I provide empirical evidence for this by testing for violations of the random assignment assumption. Table B.1 displays the p-value of a joint F-test resulting from a multinomial logit of assigned cohort on officer characteristics for each of the three exams separately. The main sample (Exam 2010) cohorts have the highest p-values and are far from statistically significant even when including additional officer factors such



Note: Figure displays the share of cohorts with more than 10 starting members that are Black (Panel A) and speak Spanish (Panel B), and the lowest starting age (Panel C) by the cohort start date, from 2009 to 2016. Exam denotes the time period during which the cohorts started and assumes cohorts in the same period were in the same test pool.

Figure B.1: Composition of Cohorts by Start Date

Main Sample	Exam	Controls	Multinomial Logit P-Value	N Recruits	N Cohorts
No	2006	Minority, Gender, Start Age	0.155	360	5
No	2006	+ Military, Spanish, High Edu.	0.000	360	5
Yes	2010	Minority, Gender, Start Age	0.860	1135	21
Yes	2010	+ Military, Spanish, High Edu.	0.444	1135	21
No	2013	Minority, Gender, Start Age	0.307	1253	17
No	2013	+ Military, Spanish, High Edu.	0.081	1253	17

Note: Table reports the p-value of the joint F-test on the coefficients of a multinomial logit regressing assigned cohort on officer characteristics for each exam period for two sets of controls. The limited controls include the officer being a minority, start age, and gender; the second set of controls adds if they were in the military, if they speak Spanish, and if they have a Bachelors degree or higher.

Table B.1: Multinomial Logit for Cohort Assignment

as education, military status, and Spanish language ability. The Exam 2013 cohorts have a non-significant p-value with the main controls (race, gender, start age); however, it becomes marginally significant when additional controls are introduced.[3] There are only 5 Exam 2006 cohorts, yet the p-value is not statistically significant with the limited controls– adding additional controls produces a statistically significant p-value, however there are more predictors than potential cohorts. These results indicate that officer characteristics are not predictive of assigned cohort, particularly in the main sample.

For additional evidence of random assignment, I test if officer characteristics are significantly associated with cohort composition. Table B.2 displays the results of regressing an officer's cohort characteristics on their individual characteristics. Columns (1)-(3) focus on the main sample with baseline controls (minority, gender, start age); Columns (4)-(6) also focus on the main sample but include additional variables (military status, Spanish ability, and education level); Columns (7)-(9) repeat the analysis of Column (4)-(6) with the full sample. Based on Columns (1)-(3), the baseline controls explain very little of the variation in cohort composition and all statistically significant coefficients are economically insignificant: being Male has the largest statistically significant effect, but it implies that being male is associated with only a 0.21% increase in average cohort age relative to the mean. As expected, being male is negatively associated with cohort share male, just as being a minority is negatively associated with cohort share minority– since cohort shares exclude

[3]The CPD's demographic data often combines race and ethnicity into a single variable. For expositional purposes and due to the data used, I will refer to 'Hispanic' as a distinct racial group.

the officer in question, it reduces the pool of officers with that characteristic, as noted in [300]. The additional officer characteristics in Columns (4)-(6) are also generally statistically and economically insignificant. Using the full sample in Columns (7)-(9) supports the lack of economic or statistical significance in the full sample as well.

Given that cohorts begin successively and not all at the same time, it is likely some amount of selection out of the academy by officers who give up, find other jobs, are no longer eligible (too old, moved out of Chicago, could not pass the physical exam, etc.)– though this may be limited in Chicago relative to other police departments as the CPD is highly oversubscribed and a well-paying department. We want to ensure that this delayed entrance and selection does not significantly alter the composition of recruits. Column (10) of Table B.2 regresses when the officer started at the academy (in years since 2009) on officer characteristics for the full sample. An officer being a minority, in the military, a Spanish-speaker, or highly educated (Bachelors or higher) have no statistically significant effect on when they start at the academy. Unsurprisingly, officer start age is statistically significant and positively associated with start date, but as the coefficient is less than 1 it implies there is censoring (at 40) and likely some selection out for aging officers. A recruit being male is associated with an earlier start date, implying male applicants may exit the pool quicker than female applicants, though being male is only associated with starting the academy about 0.055 years before female applicants, which equates to about 20 days. While attrition from the sample pool is almost certain, the evidence presented indicates it is not likely to significantly impact the composition of cohorts or be associated with differences in officer unobservables. Furthermore, applicants wait over a year before the first cohort is called in, meaning the least committed applicants likely select out once they receive their numbers and understand where they are in the pool.

Finally, there are two points relating to random assignment and identification that should be made. First, identification of peer composition's effect on outcomes is possible thanks to variation in cohort composition differing due to random draws. If there were, for example, only two large cohorts from an exam, the variation is expected to be small. In the main sample, there are 21

	Main Sample						Full Sample			
	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Cohort Mean Age	Cohort Share Minority	Cohort Share Male	Start Date (years)
Minority	0.014	0.002	−0.005	−0.003	0.001	-0.008^{**}	0.022	0.000	-0.005^{**}	0.028
	(0.023)	(0.002)	(0.003)	(0.021)	(0.002)	(0.003)	(0.018)	(0.001)	(0.002)	(0.022)
Male	0.064^{**}	-0.009^{***}	0.005	0.063^{**}	-0.009^{***}	0.005	0.001	−0.002	0.001	-0.055^{**}
	(0.028)	(0.003)	(0.004)	(0.027)	(0.003)	(0.003)	(0.020)	(0.004)	(0.002)	(0.027)
Start Age	0.000	0.000^{*}	0.000	0.000	0.001^{*}	0.000	0.002	−0.000	0.001^{*}	0.010^{***}
	(0.004)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
Military				−0.078	0.003	−0.005	0.097	−0.003	0.003	0.194
				(0.060)	(0.005)	(0.009)	(0.097)	(0.010)	(0.007)	(0.169)
spanish				0.076	0.005	0.012^{**}	0.013	0.004	0.007	−0.003
				(0.058)	(0.005)	(0.005)	(0.048)	(0.003)	(0.005)	(0.062)
High Edu				−0.028	0.001	−0.004	−0.039	0.004^{*}	−0.003	−0.024
				(0.040)	(0.002)	(0.003)	(0.025)	(0.002)	(0.003)	(0.032)
Exam 2010							0.489^{**}	0.022	-0.043^{***}	2.875^{***}
							(0.245)	(0.034)	(0.015)	(0.293)
Exam 2013							-1.420^{***}	−0.010	−0.021	5.496^{***}
							(0.255)	(0.035)	(0.018)	(0.331)
Intercept	29.976^{***}	0.792^{***}	0.482^{***}	30.067^{***}	0.788^{***}	0.489^{***}	29.412^{***}	0.781^{***}	0.521^{***}	0.945^{***}
	(0.208)	(0.016)	(0.019)	(0.203)	(0.017)	(0.021)	(0.218)	(0.039)	(0.015)	(0.349)
R^2	0.002	0.014	0.003	0.006	0.016	0.009	0.725	0.093	0.062	0.894
Adj. R^2	−0.000	0.012	0.001	0.001	0.011	0.003	0.724	0.090	0.060	0.893
Num. obs.	1135	1135	1135	1135	1135	1135	2748	2748	2748	2748

Note: Table display results for balance regression tests. Each column displays the coefficients of officer characteristics on their cohort composition (Columns (1)-(9)) and start date (Column (10)) for main sample (Columns (1)-(6)) and full sample (Columns (7)-(10)) officers. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level (unless otherwise specified) are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2: Balance Regressions

cohorts, and as expected, the variation in cohort composition is larger in cohorts that are smaller: the correlation between cohort size and the absolute difference between cohort share minority and the mean cohort share minority is -0.4. Second, there may be a concern that cohort size is related to unobservable trends or departmental demands that influence officer outcomes. For example, the department may be expecting a low-crime period, and thus draw in fewer officers (a smaller cohort), which may be an issue for estimates relying on cohort composition variation. However, this is unlikely because the CPD must draw in applicants more than 1.5 years before they can actually begin as police officers, and these decisions take into account not only (likely imprecise) crime projections but also city budgets, staffing demands, and various other concerns.

B.1.3 Attrition

If the likelihood of attrition from the sample is impacted by diversity of one's cohort, then results in my estimation may be driven by selection bias rather than actual peer effects. In Table B.3, I present results for logistic regressions where each outcome is a form of attrition for officers in the main sample cohorts (Columns (1)-(5)) and full sample (Column (6)-(10)).

The outcome of Column (1) and (6) is whether the officer is not in the daily assignment (AA) data (52 main sample recruits). The outcome of Column (2) and (7) is outcome is whether the officer, conditional on being in the AA data, spent too much or too little time in the academy or probationary period (113 main sample recruits). The outcome of Column (3) and (8) is is whether the recruit was not in the final AA data, conditional on the previous two restrictions, meaning they were not matched to the salary and rank data as a police officer (24 main sample recruits). The outcome of Column (4) and (9) is any form of attrition across all recruits, including whether fixed effects could be recovered. The outcome of Columns (5) and (10) is any attrition from the sample (as in (4) and (9)), and also does not appear in the training data (data on specific courses the recruits attended). As displayed across all columns for the main sample, there is no statistically significant predictor of any form of attrition with respect to cohort composition (neither cohort diversity nor mean age), thus it is unlikely that attrition driven selection is driving my results. For the full sample, only peer age is ever statistically significant (Column (8)).

Another form of attrition is sample attrition after the recruits exit the academy, become full officers, and are present in the assignment data, e.g. cohort diversity being related to when officers choose to retire or exit the assignment data. While this may cause some officers to be more represented in the sample than others, the fixed effects recovered for the main sample are based on over 100 observations for almost all officers (96.6% of main sample recruits). I test for sample-exiting attrition in Table B.4. Column (1) study the relationship between cohort share minority and officer's number of observations in the assignment data used to estimate fixed effects, showing no significant relationship. Column (2) shows that cohort diversity has no effect on whether the officer exists in the salary and unit history data (which contains officers not in the assignment data) at the end of 2018. Column (3) shows cohort diversity has no significant effect on likelihood of being promoted by the end of 2018 and that average peer age has a statistically significant but economically small effect. Column (4) shows the likelihood of being in a non-geographic unit at the end of 2018 is not significant impacted by cohort composition.

	Main Sample					Full Sample				
	Not in AA	Training Time Violation	Not in Final AA	Any Attrition	Not in Trainings	Not in AA	Training Time Violation	Not in Final AA	Any Attrition	Not in Trainings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cohort Share Minority	−1.06	1.31	0.71	0.22	−1.48	0.19	1.46	−0.22	0.22	0.01
	(3.14)	(3.81)	(3.49)	(2.77)	(3.00)	(1.95)	(3.06)	(2.60)	(2.73)	(2.12)
Cohort Mean Age	0.51	−0.03	−0.72	0.05	−0.18	0.09	−0.02	-0.63^{**}	0.05	−0.32
	(0.37)	(0.29)	(0.41)	(0.23)	(0.32)	(0.18)	(0.29)	(0.22)	(0.23)	(0.23)
Black	−0.24	0.19	-15.83^{***}	0.03	0.26	−0.06	0.24	−0.14	0.03	0.25
	(0.47)	(0.41)	(0.40)	(0.34)	(0.30)	(0.22)	(0.23)	(0.43)	(0.33)	(0.15)
Hispanic	0.43	0.19	−0.03	0.25	0.28^{*}	0.30	0.25	0.10	0.25	0.25^{**}
	(0.36)	(0.24)	(0.55)	(0.19)	(0.14)	(0.20)	(0.15)	(0.33)	(0.19)	(0.09)
Asian/Native American	−0.68	0.47	-16.04^{***}	−0.01	0.35	−0.82	0.07	−0.86	−0.01	−0.10
	(0.90)	(0.45)	(0.43)	(0.40)	(0.32)	(0.70)	(0.37)	(0.97)	(0.39)	(0.23)
Male	−0.05	−0.40	0.29	−0.28	−0.19	−0.21	−0.13	0.00	−0.28	−0.14
	(0.39)	(0.24)	(0.59)	(0.23)	(0.19)	(0.25)	(0.18)	(0.34)	(0.23)	(0.12)
Start Age	0.02	−0.02	-0.13^{*}	−0.02	-0.04^{*}	−0.00	−0.01	-0.10^{**}	−0.02	-0.03^{*}
	(0.02)	(0.03)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
Cohort Size	0.00	−0.03	−0.00	−0.02	−0.01	−0.00	-0.02^{*}	−0.00	−0.02	0.00
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Exam 2010						−0.43	−0.64	0.33	-22.64^{***}	−0.66
						(0.37)	(0.35)	(0.48)	(0.62)	(0.40)
Exam 2013						−0.38	-1.22^{**}	-1.21^{*}	0.02	-2.26^{***}
						(0.44)	(0.41)	(0.56)	(0.58)	(0.56)
Intercept	−18.97	0.58	21.20	−1.45	6.66	−5.17	−0.41	17.91^{**}	21.19^{***}	10.17
	(11.20)	(7.68)	(13.34)	(6.07)	(10.13)	(5.41)	(7.30)	(6.86)	(5.91)	(6.60)
AIC	433.52	710.84	226.46	1039.10	1347.12	1009.95	1439.73	538.15	1043.10	2826.49
Log Likelihood	−207.76	−346.42	−104.23	−510.55	−664.56	−493.97	−708.86	−258.07	−510.55	−1402.25
Num. obs.	1135	1083	970	1135	1135	2748	2626	2409	2748	2748

Note: Table display the logistic regression estimates of cohort and officer observables on officer attrition for various reasons from the main sample (July 2012 - May 2014 cohorts). The dependent variables for Columns (1)-(3) are: (1) whether or not the officer is not matched in the assignment data; (2) whether or not the officer is dropped due to spending too much or too little time in the academy or probationary period; (3) whether or not the officer is not in the final assignment data, meaning they were matched in the salary and unit history data and spent some time as a D1 officer in units 1-25; (4) attrition for any of the listed reasons or if no fixed effects could be recovered due to too few observations. Columns (2)-(3) are estimated on the sample of recruits which were not dropped due to the previous column's reason, and Column (4) is estimated on all initial sample recruits. Column (5) is whether or not the officer appears in the training cohort sample, meaning no attrition (Column (4)) and is in the data on trainings/classes. Columns (6)-(10) replicate (1)-(5) for the full sample. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table B.3: Attrition from Sample

	N. Obs in Data	Exit Data	Promoted at End	Specialized Unit at End
	(1)	(2)	(3)	(4)
Cohort Share Minority	−315.73	−0.03	−0.06	−0.07
	(256.54)	(0.13)	(0.12)	(0.11)
Cohort Mean Age	-52.42^{*}	0.02	-0.03^{*}	−0.02
	(24.14)	(0.01)	(0.01)	(0.01)
Black	−28.94	−0.01	-0.04^{*}	-0.07^{***}
	(15.56)	(0.02)	(0.02)	(0.02)
Hispanic	16.39	0.01	−0.03	-0.05^{***}
	(12.10)	(0.01)	(0.02)	(0.01)
Asian/Native American	24.21	0.03	−0.04	−0.01
	(23.06)	(0.03)	(0.03)	(0.03)
Male	99.24^{***}	0.02	0.01	0.01
	(20.26)	(0.02)	(0.01)	(0.01)
Start Age	2.03	0.00^{**}	−0.00	-0.00^{**}
	(1.40)	(0.00)	(0.00)	(0.00)
Cohort Size	−0.77	0.00	0.00	0.00^{*}
	(0.52)	(0.00)	(0.00)	(0.00)
Exam 2010	-287.01^{***}	0.06^{*}	−0.04	-0.04^{**}
	(43.96)	(0.02)	(0.02)	(0.02)
Exam 2013	-634.17^{***}	0.16^{***}	-0.26^{***}	-0.21^{***}
	(51.18)	(0.03)	(0.02)	(0.02)
Intercept	2461.07^{***}	0.11	1.04^{**}	0.87^{**}
	(658.18)	(0.34)	(0.34)	(0.29)
Mean Dep. Var	446.17	0.89	0.11	0.07
R^2	0.42	0.03	0.09	0.09
Adj. R^2	0.42	0.02	0.08	0.08
Num. obs.	2497	2748	2441	2441

Note: Table display the linear regression estimates of cohort and officer observables on officer observations and other measures of attrition for the full sample. The dependent variables are the officer's number of observations (shifts) used to estimate fixed effects in the daily panel data (Column (1)), whether or not the officer is in the salary and unit history data which contains non-D1 officers and units outside of the assignment data (Column (2)), whether the officer has been promoted by the end of 2018 (Column (3)), whether the officer is in a specialized unit at the end of 2018 (Column (4)). Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table B.4: Attrition out of Sample

B.1.4 Confounding Assignments

While cohort composition is not related to a new officer's likelihood of attrition, cohort composition may influence where an officer works. The CPD's unit (district) assignment process is seniority based, with new recruits being assigned based on departmental demand, avoiding very long commutes for officers.[4] and possibly a desire to have officers reflect the communities they police. As a result, it is possible that assignment is influenced by one's cohort's composition, since if Black officers are more likely to be initially placed in Black areas (for example, if the department attempts to place Black officers in districts with Black civilians or those officers are more likely to live close by), having more Black cohort-mates may reduce another officer's probability of being placed in a majority-Black district. In this subsection, I explore the effect of peer diversity on assignment.

How departmental need and initial assignments are determined is not clear, and there does not seem to be a significant correlation between a recent decline in the number of officers in a unit and the share of new recruits serving in that unit. In the data, however, there is a clear relationship between the race of officers serving in a district and that of the district's population. Furthermore, as officers gain seniority and the ability to leave high-crime (high minority population) districts, they do so, meaning the districts in need of officers are higher crime areas.

Table B.5 displays the characteristics of the average unit (district) in which officers work for officers in the full panel. These simple regressions explain some amount of variation in officer assignments. Based on the results, new officers are much more likely to be placed in high crime areas both during and after their probationary periods—this may be partially explained by those units demanding the most officers and higher seniority officers transferring to less dangerous areas. As expected, assignments are influenced by officer race: being Black increases the share of Black civilians in the average district in which an officer works by over 8 percentage points. Similarly, there is a clear relationship between officers being Black and their districts having higher crime and lower income. These results are, however, for all officers in the panel which means the estimates

[4] This information is based on conversations with a retired officer.

are noisy due to significant changes in crime and recruit composition over time.

Focusing on the main sample officers, Table B.6 displays regression estimates predicting the characteristics of the average unit a new officer serves in as a full officer. Notably, having more Black and Hispanic peers in one's academy cohort decreases the Black population share of the average district in which an officer works. As is evidenced from the table, Black officers are more likely to serve in Black districts, thus it is likely the case that having more Black officers in one's cohort reduces the probability of officers to be placed in Black districts.

However, this confounding assignment is unlikely to significantly bias estimates because new officers work in high crime areas with higher Black populations regardless of an officer's race, as shown in Table B.5. Despite this, main sample officers in the work in all of the 22 districts during the 2013-2018 period and no single district makes up more than 11% of assignments. Furthermore, it is evident from the very small amount of variation explained by observables in Table B.6 that there is likely much CPD-level demand choices made regardless of cohort observables and the actual influence of such observables are economically small. For example, a 5pp increase in cohort share Black only decreases a recruit's average district's Black population share by about 6% of the baseline mean. Given this, it is unlikely that the influence of cohort composition on new officers' assignments will significantly bias results, and controlling extensively for working environment, as in Section 2.3, should remove this bias completely.

B.1.5 Working Peers and Instructors

Another concern is that officers exposed to higher amounts of minorities in their cohorts may end up working with more minorities in the future. If contemporaneous peers influence arresting decisions, then the effect of academy peers may be capturing the selection of future minority peers and their influence. A similar concern is that academy diversity may influence the composition of one's field training officers. To test these, I regress the average composition of main sample officers' contemporaneous peers and field training officers on their cohort composition. Table B.7 displays these results.

	Violent Crime	Median Income	Share Black Pop.	Share Hispanic Pop.
	(1)	(2)	(3)	(4)
Probationary	2.60***	−8169.67***	0.15***	−0.02***
	(0.12)	(304.27)	(0.01)	(0.00)
Recruit	4.70***	−11446.91***	0.23***	−0.04***
	(0.16)	(358.97)	(0.01)	(0.01)
Female	−0.20	560.99	−0.01	−0.01
	(0.12)	(345.19)	(0.01)	(0.00)
Black	3.26***	−5486.64***	0.26***	−0.13***
	(0.14)	(404.67)	(0.01)	(0.00)
Hispanic	0.61***	−1814.42***	0.02**	0.02***
	(0.13)	(333.13)	(0.01)	(0.00)
Other Race	−0.52*	1156.56	−0.04***	−0.01
	(0.27)	(720.41)	(0.01)	(0.01)
Intercept	8.97***	50594.28***	0.35***	0.26***
	(0.09)	(270.61)	(0.00)	(0.00)
R^2	0.12	0.09	0.15	0.08
Num. obs.	12904	12904	12904	12904

Note: Table displays the linear regression estimates of officer characteristics on the average characteristics of the districts in which they work. Population and income are determined based on 2010 Census estimates. Violent crime rates are violent crimes in a month, based on Chicago City Data Portal crime data, per 10,000 population. The coefficients Recruit and Probationary are indicators for whether or not the officer is a new officer in their post-probationary period or a new officer in their probationary period, respectively. Robust standard errors are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.5: Characteristics of Average Working District

	Violent Crime	Median Income	Share Pop. White	Share Pop. Black	Share Pop. Hispanic
	(1)	(2)	(3)	(4)	(5)
Cohort Share Black	−11.12	10408.85	0.17	-0.92^{**}	0.70^{**}
	(8.44)	(9381.91)	(0.11)	(0.36)	(0.30)
Cohort Share Hispanic	−0.72	−3074.69	−0.01	-0.43^{*}	0.45^{**}
	(4.72)	(4902.40)	(0.07)	(0.23)	(0.18)
Cohort Share Other	7.20	54.20	−0.01	0.27	−0.22
	(8.98)	(7795.52)	(0.10)	(0.44)	(0.35)
Cohort Mean Age	−0.47	545.09	0.01	−0.00	−0.01
	(0.48)	(588.76)	(0.01)	(0.03)	(0.02)
Black	0.04	263.77	−0.01	0.09^{***}	-0.08^{***}
	(0.63)	(1019.01)	(0.01)	(0.03)	(0.02)
Hispanic	−0.18	76.50	0.00	−0.01	0.01
	(0.50)	(653.48)	(0.01)	(0.02)	(0.02)
Asian/Native American	0.09	550.50	0.05^{**}	−0.07	−0.01
	(1.60)	(1805.14)	(0.03)	(0.06)	(0.04)
Male	−0.11	373.01	0.01	0.00	−0.01
	(0.62)	(912.47)	(0.01)	(0.03)	(0.02)
Start Age	-0.13^{**}	106.82	0.00	-0.01^{**}	0.00^{***}
	(0.05)	(91.07)	(0.00)	(0.00)	(0.00)
Cohort Size	-0.04^{**}	22.48	−0.00	−0.00	0.00^{*}
	(0.02)	(18.07)	(0.00)	(0.00)	(0.00)
Intercept	37.60^{**}	12463.55	−0.11	1.20	−0.07
	(14.58)	(18605.44)	(0.19)	(0.73)	(0.60)
Mean Outcome	15.49	34203.262	0.085	0.7	0.184
R^2	0.02	0.01	0.01	0.04	0.05
Num. obs.	940	940	940	940	940

Note: Table displays the linear regression estimates of recruit and cohort characteristics on the average characteristics of the districts in which they work after their probationary period. Population and income are determined based on 2010 Census estimates. Violent crime rates are violent crimes in a month, based on Chicago City Data Portal crime data, per 10,000 population. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.6: Characteristics of Average District in Post Probationary Period

In Columns (1) to (4), I use two main peer groupings, officers assigned to the same shift and watch number and the same sector (a subset of a district composed of multiple beats) or same beat, with Columns (1) and (3) being during the officer's probationary period and (2) and (4) being during the officer's time as a full officer. The dependent variable in Column (5) is the share of an officer's field training officers who are white during their probationary period. Relative to the mean white share across each group and officer type (probationary and full), the relationship between an officer's cohort share minority and their future peers' and FTOs' share white is not economically significant, and only one is statistically significant. The effect size with the largest magnitude (Column (3)) indicates that a 10pp increase in cohort share minority leads to a 0.016 decrease in the peer share white, which is a 3.4% decrease relative to the mean. The small and noisy, but consistently negative, relationship between cohort share minority and the share of white working peers is likely an artifact of this model not controlling for unit assignment: more minority peers slightly crowd out positions in high Black areas which also have slightly fewer white officers. But, as discussed in Appendix B.1.4, this effect is minor and almost all new officers go to high crime and high share Black districts. Given this, it is unlikely that the effect of future peer or training officer composition is driving the effects, and much of these small and noisy effects can be explained by the weak influence of cohort diversity on unit assignment. Furthermore, the extensive assignment controls discussed in Section 2.3 will absorb a significant amount of the minor differences in working peers as well.

While not studied in this paper, an officer's partner working with them in the same car or same beat on a specific day likely also influences their behavior. New officers are generally required to be placed with more experienced ones, so it is unlikely that cohort-mates work together. Furthermore, the working day rotation system creates a relatively chaotic environment during an officer's first few years in terms of partners. Essentially, who a new officer is able to work with is determined by the rotating day-off-group calendar, officers taking furlough days, and the fact that the units new officers are placed (high violent crime) are those that more senior officers tend to leave when vacancies are available elsewhere. So, for newer officers, who have the least choice in their shift, unit, and day-off-group, as well as no prior experience in the district (and thus no potential for having an established beat assignment), working-partners rotate frequently. Partners during these early years (focused on in this study) likely involve little endogenous selection relative to later on in their careers.

B.1.6 Shrinkage Estimates

The individual fixed effects recovered by equation (2.2) are based upon finite observations of officers. This means that each estimated fixed effect will have some error associated with it, and it is crucial that measurement error is not driving the results. A common procedure to correct for this when using individual fixed effects, popularized by the teacher value added literature ([301]), is to do an empirical Bayes shrinkage procedure (based on [302]). The idea is to shrink estimates (officer fixed effects) toward a prior mean based on how noisy the estimate is (high noise leads to a larger reduction in the estimate). Here, I will construct a shrunken estimate for each officer fixed effect ($\hat{\theta}_i$) based on how noisy it is ($Var[\hat{\theta}_i] = se(\theta_i)^2$) relative to the variance in the distribution of all fixed effects ($Var[\hat{\theta}] = \frac{1}{N}\sum_i^N \hat{\theta}_i{}^2$):

$$\hat{\theta}_i^{shrunken} = \hat{\theta}_i * \frac{Var[\hat{\theta}]}{Var[\theta_i] + Var[\hat{\theta}]}.$$

This is derived from the posterior mean of normal distribution with prior mean equal to zero being:

$$\theta_i^n = \bar{\theta}_i \frac{\sigma^2}{\sigma^2 + \frac{\sigma_i^2}{n}},$$

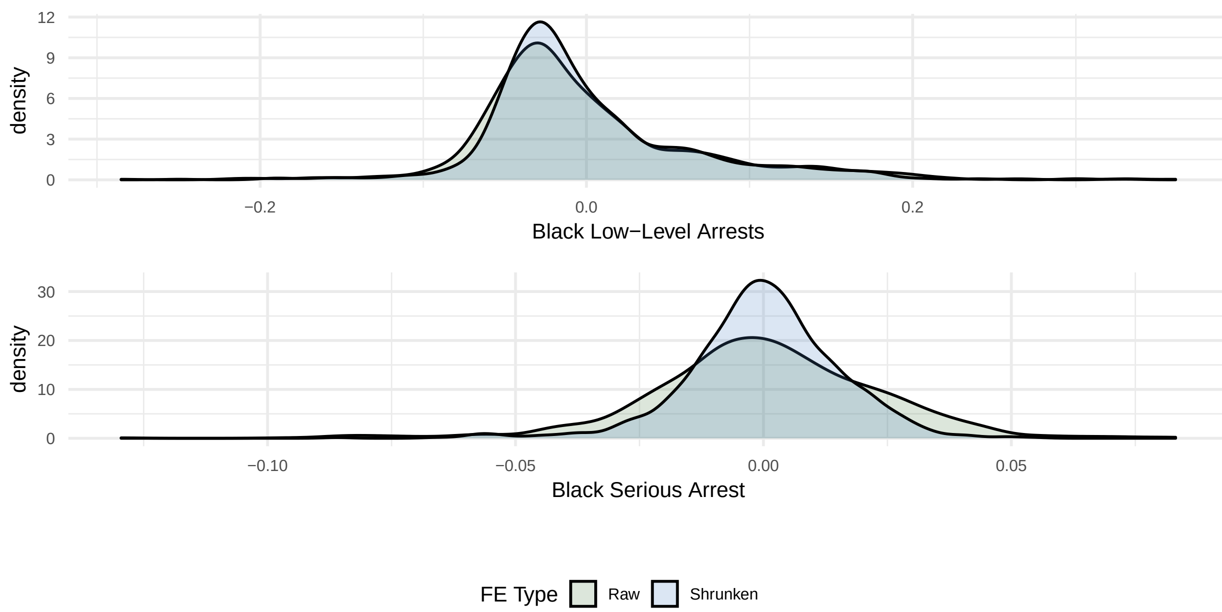
where θ_i is drawn from a $N(0, \sigma^2)$ and each observation of $\theta_i^t = \theta_i + \epsilon_i^t$, where $\epsilon_i^t \sim N(0, \sigma_i^2)$. $\bar{\theta}_i$ can be seen as the fixed effect estimate, $\frac{\sigma_i^2}{n}$ is the $se(\bar{\theta}_i)^2$, and σ^2 as the variance across estimates of θ_i's.[5]

[5] A key advantage of the 'lfe' R package is its ability to estimate standard errors for fixed effects via bootstrapping. However, as discussed in [303], bootstrapped standard errors tend to be conservative; if this is the case in my environment, I may be 'over'-shrinking because overly conservative (large) standard errors will lead to smaller rescaling factors. However, this is unlikely to be a concern as the results are consistent with raw fixed effects as well as post-shrinking fixed effects.

	Average Sector Share White		Average Beat Share White		FTO Share White
	Probationary	Full	Probationary	Full	Probationary
	(1)	(2)	(3)	(4)	(5)
Cohort Share Minority	−0.01	−0.13	-0.16^{**}	−0.09	0.06
	(0.07)	(0.11)	(0.08)	(0.17)	(0.15)
Cohort Mean Age	0.01^{*}	0.01^{**}	−0.00	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Black	-0.10^{***}	-0.15^{***}	-0.12^{***}	-0.29^{***}	-0.14^{***}
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Hispanic	-0.03^{***}	-0.06^{***}	-0.05^{***}	-0.16^{***}	-0.04^{*}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Asian/Native American	0.04^{**}	−0.02	0.02	-0.11^{***}	0.05
	(0.02)	(0.02)	(0.04)	(0.03)	(0.05)
Male	−0.01	0.03^{***}	−0.00	0.05^{***}	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Start Age	0.00	-0.00^{**}	−0.00	-0.00^{**}	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cohort Size	−0.00	−0.00	−0.00	-0.00^{*}	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	0.22	0.22	0.68^{***}	1.05^{**}	0.94^{**}
	(0.18)	(0.24)	(0.25)	(0.44)	(0.46)
Mean Dep.	0.48	0.48	0.46	0.48	0.43
R^2	0.08	0.20	0.06	0.25	0.03
Num. obs.	940	940	940	940	940

Note: Table displays the linear regression estimates of main sample officer characteristics on their working peers and training officers during their probationary (odd columns) and post-probationary (even columns) periods. The dependent variable in Columns (1) and (2) is the share of an officer's peers who are white and working the same day, shift, and sector. The dependent variable in Columns (3) and (4) is the share of an officer's peers who are white and working the same day, shift, and beat number. The dependent variable in Column (5) is the share of an officer's field training officers who are white. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.7: Average Characteristics of Peers and Training Officers



Note: The figure compares the centered distributions of main sample officer fixed effects for arrests of Blacks for low−level and serious crimes, recovered from equation (2). Raw fixed effects are unaltered, while shrunken fixed effects were subject to the Bayesian shrinkage procedure described in Appendix A.6.

Figure B.2: Distributions of Raw and Shrunken Main Sample Fixed Effects

As expected, due to the relatively large number of observations per officer in my sample (the median main sample officer has 575 observations and the median full sample officer has 401 observations), the shrunken fixed effects are similar to those of the main results. Figure B.2 displays the distribution of raw and shrunken fixed effects for arresting Blacks for low-level and serious crimes. For low-level arrests, which are more common, the raw and shrunken distributions are more similar.

B.1.7 Randomization-Based Inference

Randomization inference (or randomization-based inference) allows us to construct an empirical distribution of coefficients under the null hypothesis, that peers have no effect on officer fixed effects. This is preferable to traditional asymptotic inference in which the error in estimates is a result of sampling error because in such environments, there is no sampling error; e.g., the sample of CPD recruits between 2009 and 2016 is the population. Such methods have their origin in [304], wherein one wants to test to see if they can reject the 'sharp' null hypothesis that the treatment has no effect on the outcome of interest, and much of this section will follow [305]. Let us generalize equation (2.3) (removing superscript k for simplicity) as a potential outcomes function:

$$\theta_i(P_i = \overline{X}_{c(i)}) = \alpha_{p(i)} + \pi_1 \overline{X}_{c(i)} + \pi_2 X_i + v_i$$

Then the potential outcomes function for an individual, θ_i, takes in a value for i's peer composition P and tells us what the individual's fixed effect for arrests would be had they had peer composition P in the academy. As discussed in [305], under a sharp null hypothesis of no effect, given some treatment assignment P' and the realized outcomes for that specific assignment $\theta_i(P')$, one can infer the value of the outcome at any other treatment assignment. Essentially if under the null that $\pi_1 = 0$, then $\theta_i(P) = \theta_i(P') \forall P, P' \in \mathbb{P}$ where P' is any possible peer composition and $\mathbb{P}$ is the space of all possible treatment (peer) assignments. The intuition is that if the true peer effect is zero ($\pi_1 = 0$), then it should not matter what treatment (peer composition) is assigned.

Now, we can test this null hypothesis. We can generate test statistics based on the distribution of estimated treatment effects (π_1^r), the 'randomization distribution', when the treatment status is randomly assigned. With this distribution of estimated peer effects under the randomized treatments, we compare the estimate from our actual data ($\hat{\pi}_1$) to the randomization distribution and recover the p-value– the likelihood of finding an effect more extreme than the one estimated under the null hypothesis that treatment has no effect. Again, borrowing from [305]:

$$\textit{p-value} = Pr(|\hat{\pi}_1(\theta_i(P = \overline{X}_{c(i)}))| \geq |\pi_1^r(\theta_i(P'))|)$$

With this p-value, we can assess the likelihood that the estimate recovered from the actual data ($\hat{\pi}_1$) is consistent with the null hypothesis, that the the peer effect is null.

In practice, constructing the randomization distribution can be done in two ways. (1) (Re-assigning Treatment) Randomly re-assigning individuals to cohorts within exams and ensuring cohort sizes remain the same and thus constructing randomized treatments ($\bar{X}^r_{c^r(i)}$), then estimate:

$$\hat{\theta}_i = \alpha_{p(i)} + \pi_1^r \overline{X}^r_{c^r(i)} + \pi_2 X_i + v_i$$

Or (2) (Re-assigning Outcomes), randomly re-assigning outcomes to individuals (θ_i^r):

$$\hat{\theta}_i^r = \alpha_{p(i)} + \pi_1^r \overline{X}_{c(i)} + \pi_2^r X_i + v_i$$

In either case, this procedure can be repeated N number of times (I perform 1,000 iterations for method (1) and 5,000 iterations for method (2)) with each iteration producing an estimate of π_1^r. Then, the coefficient using the actual data, $\hat{\pi}_1$ can be compared with the distribution of $\hat{\pi}_1^r$ to obtain a p-values as discussed above. Method (1) is used in [209] and [306], while Method (2) is used in [208] and [174].[6]

Effectively, method (1) differs from (2) in that (1) randomizes only the treatment of interest while (2) also scrambles the relationship between the controls (X_i) and the outcome. In practice because of sample attrition, method (1) involves re-drawing cohorts (within exams) using recruits in the final sample and those who are dropped from it, where as method (2) randomizes the outcomes (within exams) between officers who appear in the final sample.[7] In both cases, two-sided p-values are computed by ranking the coefficient in the main results within the distribution of placebo coefficients.

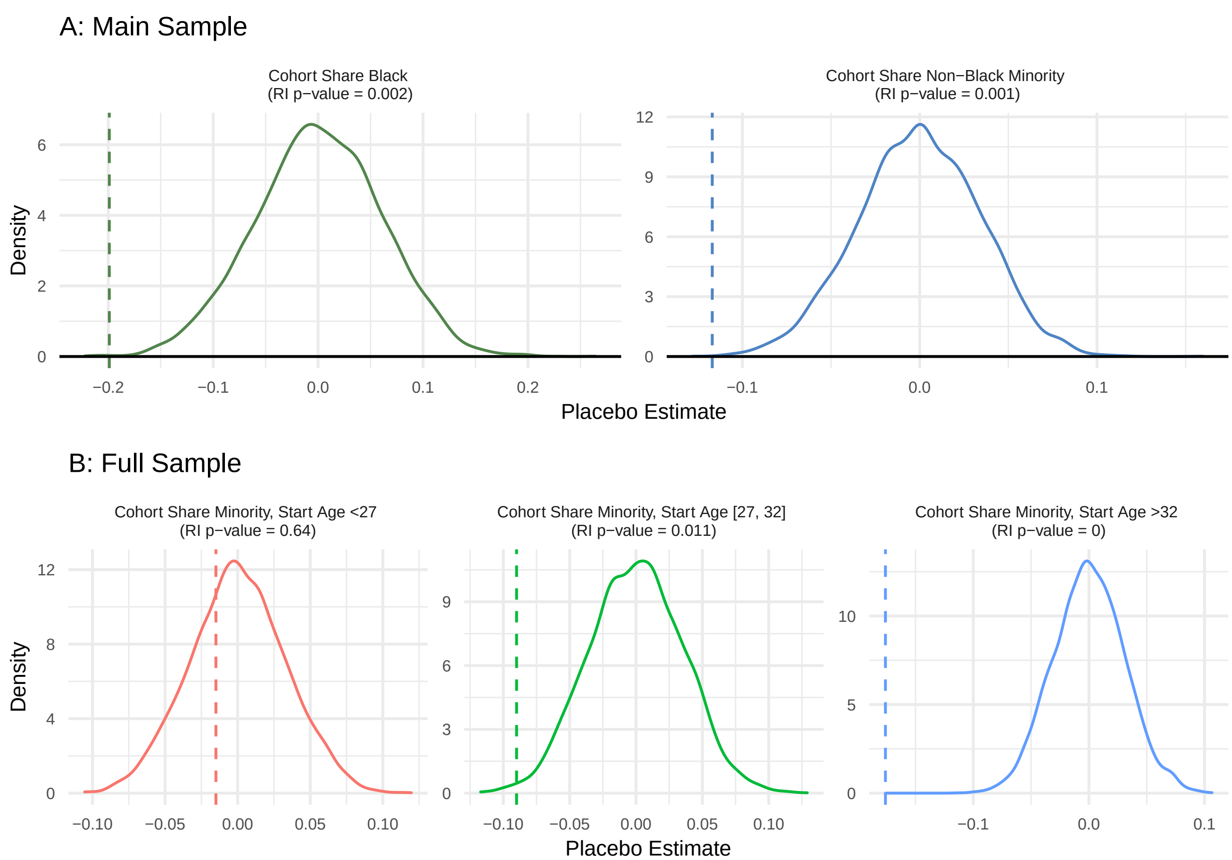
As shown in Figures 2.2 and B.3, the estimated peer effects which are reported as statistically significant in the main results are statistically significant using either method, with empirical p-values generally smaller than those in the main results.

B.1.8 Small Class Effects (Homerooms)

While many classes were composed of almost all the officers in one's cohort, smaller sub-cohort groups ("homerooms") are identifiable when restricting to classes with fewer than 30 re-

[6]The reassignment of outcomes to observations is based on the author's understanding of the replication code for [174], see lines 529 to 560 for reference in 20170069_main.do in their replication files.

[7]Both methods do take the error in the outcome (i.e., $\hat{\theta}_i$ is an estimate with measurement error) as given.



Note: Figure visualizes the distribution of coefficients estimated using 5,000 placebo rounds to conduct randomization inference, as discussed in Appendix A.7. Each placebo round involves randomly re−assigning the outcome variable to an officer within a exam period. Coefficients are the effects of cohort shares Black and non−Black minorities in the main sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on the outcome variable, shrunken officer fixed effects for arresting Blacks for low−level crimes. The dashed vertical lines correspond the coefficient estimated in the main specification (actual cohorts), and the RI p−value denotes the p−value resulting from a two−tailed test which ranks the magnitude of the actual coefficient among the magnitudes of the 5,000 placebo coefficients. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sample).

Figure B.3: Randomization Distribution of Coefficients (Reassigned Fixed Effects) cruits. I use data on individual classes the officers took while in the academy to see if recruits in small group (homeroom) composition is driving the effects of cohort composition on the outcomes. If this is the case, then it is more likely that instructor effects are a contributing factor.

The training data provided lists the set of classes each probationary officer took during their time at the academy. This includes classes on the data base access, report writing, terrorism, chemical and radioactive events, and use of force. Many classes are large containing almost all (or a large portion) of a cohort's members. A subset of courses contain fewer officers per class, meaning there is larger within-cohort variation on which cohort members attended these courses together.

I use the set of trainings during the academy that full sample officers took which had fewer than 30 officers attend and a sufficiently high share of the classes being from the same cohort. With this set of courses, I created a weighted undirected network of recruits within cohorts and use the "edge betweenness" clustering algorithm ([307]) in order to partition these networks into sub-communities of officers that had the strongest ties based on classes taken together. I refer to these sub-cohorts as homerooms.[8]

After some filters, the final sample of officers in the homerooms (also in the full sample) is 2,093 in 105 homerooms. Not all recruits are present in the final homeroom data due to matching issues and filters (89.6% of full sample officers are in the final homeroom data) and I restrict to homerooms with between 14 and 30 recruits. Due to the smaller size of these homerooms, there is much more variation in compositions. For example, there is 2.5 times more variation in cohort share minority for homerooms relative to cohorts. Nevertheless, homeroom and cohort compositions are highly correlated: For example, the correlation between cohort share Black and homeroom share Black is 0.77.

To see if homeroom composition is driving the results, I re-estimate equation (2.3) on the full sample using homeroom instead of cohort composition. The results are displayed in Table B.8. The effects of homeroom composition are similar to those for cohort composition (Columns (1)

[8]The edge betweenness clustering algorithm is implemented in the igraph package in R ([308]).

and (2)), while adding cohort fixed effects to see if variation between homerooms within cohorts has an effect produces statistically and economically insignificant results (Columns (3) and (4)). As noted in Section 2.6.2, this is not consistent with instructor effects driving the results.

Table B.8: Effect of Homeroom Diversity on Arrest Propensity

	Shrunken Arrest Propensity			
	Black Low-Level	Black Serious	Black Low-Level	Black Serious
	(1)	(2)	(3)	(4)
Homeroom Share Minority, Start Age <27	−0.021	0.006	−0.014	0.001
	(0.029)	(0.004)	(0.015)	(0.004)
Homeroom Share Minority, Start Age [27, 32]	−0.031	0.005	0.021	0.007*
	(0.026)	(0.004)	(0.015)	(0.004)
Homeroom Share Minority, Start Age >32	−0.073***	0.012***	−0.007	0.004
	(0.025)	(0.004)	(0.017)	(0.004)
Homeroom Share White, Start Age [27, 32]	−0.038*	0.001	−0.003	−0.002
	(0.022)	(0.004)	(0.014)	(0.004)
Homeroom Share White, Start Age >32	−0.042	0.007	0.018	0.007
	(0.031)	(0.005)	(0.021)	(0.006)
Controls	Full	Full	Full	Full
Cohort FE			X	X
R^2	0.433	0.141	0.488	0.165
Num. obs.	2093	2093	2093	2093

Note: Table displays the effect of homeroom composition on full sample officers' propensities to make arrest of Blacks for low-level and serious crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Homerooms are sub-cohorts constructed using individual class training data as described in Appendix A.8. Homeroom shares are computed as the leave-out mean of the officer's homeroom's initial composition. Full controls refers to the additional controls in Column (13) of Table 3 with homeroom size substituted for cohort size. Cohort FEs refer to cohort-specific fixed effects. Standard errors clustered at homeroom level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

B.2 Appendix B - Additional Figures and Tables

Exam	Date of administration	Attended	Passed	Failed
Police Entrance 1999	3/15/1999; 3/16/1999	3,967	No info available	No info available
Police Entrance 1999	1/5/2000	2,517	No info available	No info available
Police Entrance 2000	7/1/2000	2,053	No info available	No info available
Police Entrance 2000	1/4/2001	1,829	No info available	No info available
Police Entrance 2001	5/19/2001	1,923	No info available	No info available
Police Entrance 2002	1/12/2002	3,150	No info available	No info available
Police Entrance 2003	11/22/2003	3,875	No info available	No info available
Police Entrance 2004	11/20/2004	4,163	No info available	No info available
Police Entrance 2005	2/18/2006; 2/19/2006	4,061	3,338	723
Police Entrance 2006-1	6/4/2006	1,508	1,255	253
Police Entrance 2006-2	8/6/2006	1,025	863	162
Police Entrance 2006-3	11/5/2006	1,795	1,487	308
Police Entrance 2010	12/11/2010	8,621	7,689	932
Police Entrance 2010 make up	makeups: 3/12/2011; 6/11/2011; 9/25/2011; 12/3/2011; 6/2/2013; 12/1/2012; 3/9/2013	No info available	No info available	No info available
Police Entrance 2013	12/14/2013 & military makeups (6/28/2014; 12/7/2014; 6/13/2015; 12/6/2015)	14,788	12,877	1,911
Police Entrance 2016	4/16/2016 & make ups :12/3/2016; 12/4/2016	10,199	9,023	1,176
Police Entrance Spring 2017	4/1/2017-4/2/2017	8,620	7,437	1,183
Police Entrance Winter 2017	12/16/2017,12/17/2017 & makeup: 2/24/2018	7,294	6,418	876
Police Entrance Spring 2018	5/5/2018 & 5/6/2018 & makeup: 6/23/2018	4,273	3,789	484
Police Entrance Winter 2018	12/8/2018	4,433	3,964	469
Police Entrance Winter 2018 make up	3/9/2019	Hasn't occurred	N/A	N/A

Figure B.2.1: CPD Exam Information

Note: Figure displays information on CPD entrance exam information, the date of the exam and the numbers of applicants that attended, passed, and failed the exam.

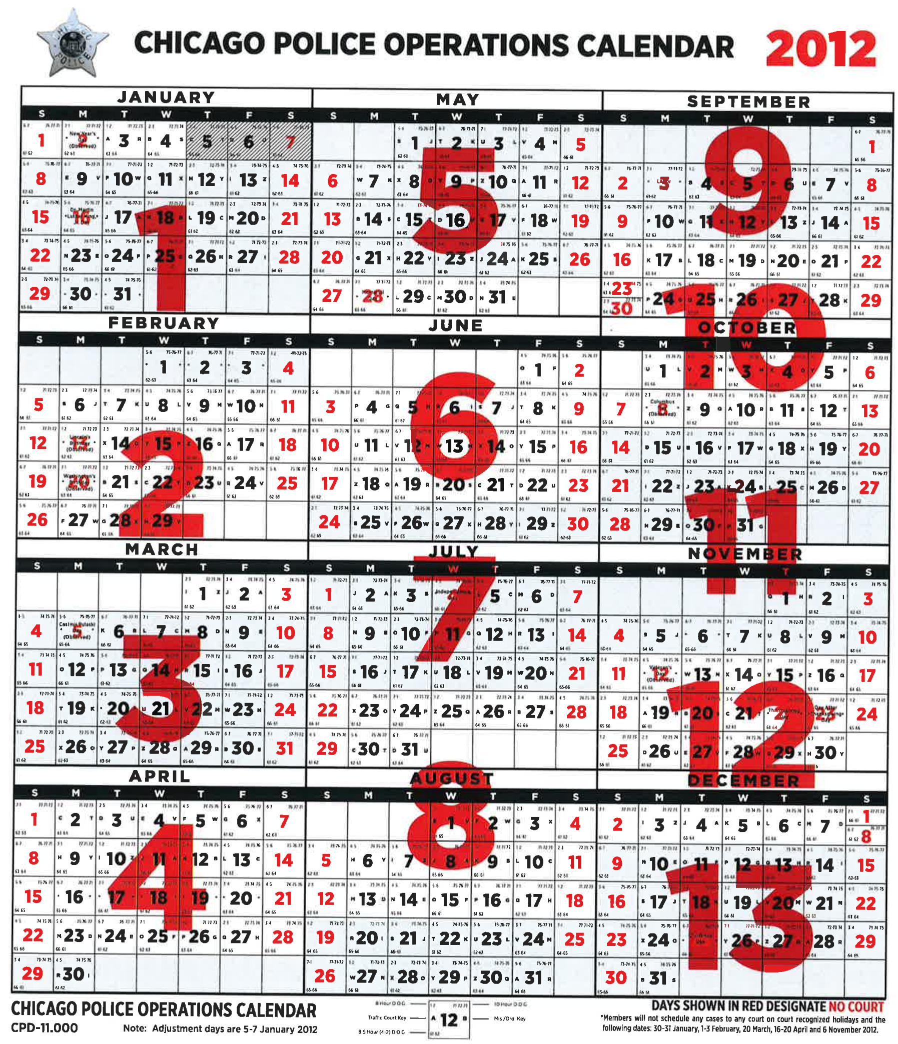
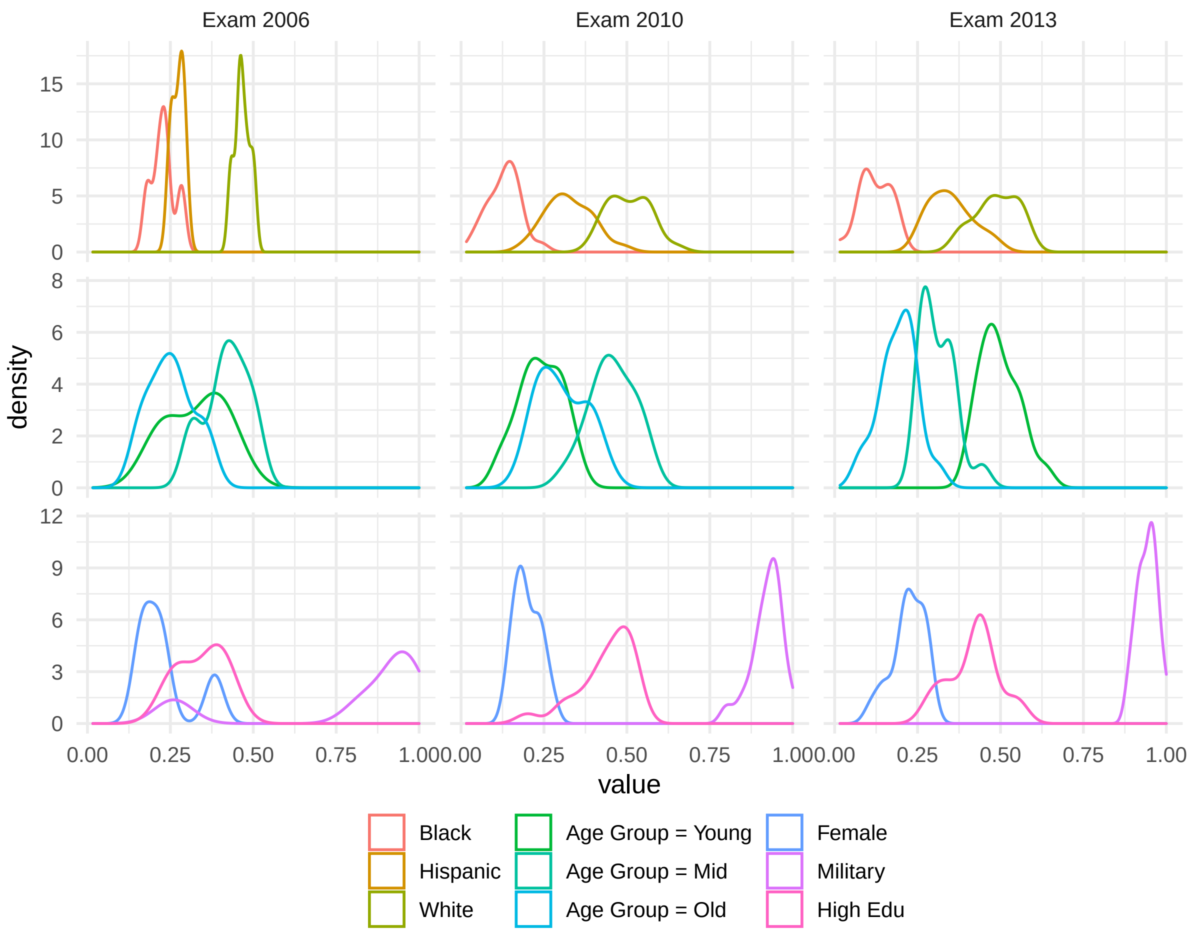


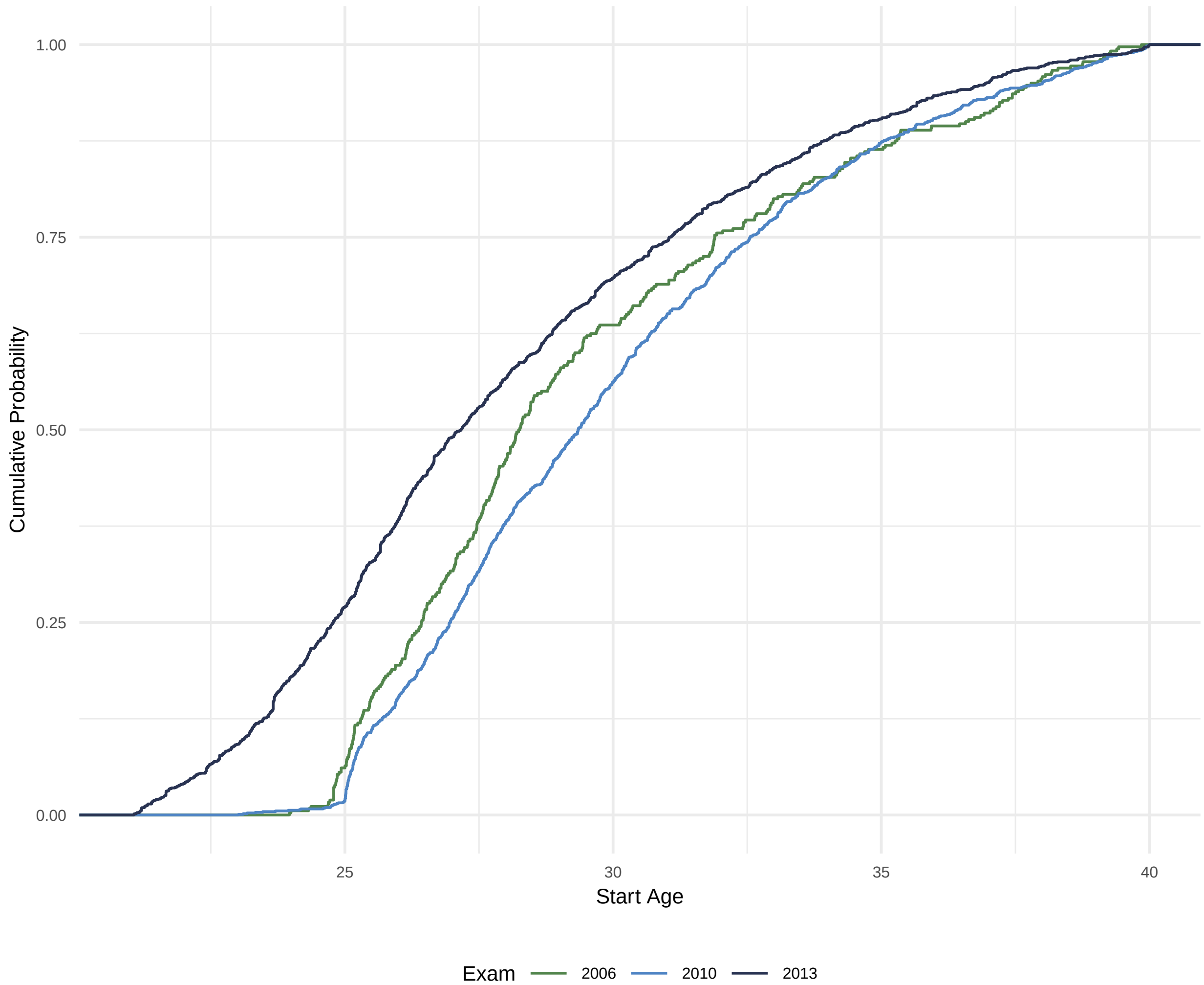
Figure B.2.2: CPD Operations Calendar (2012)

Note: Figure displays an example of the CPD operations calendar for the year 2012.



Note: Figure displays the distributions of cohort compositions for Exam periods 2010 (main sample), 2006, and 2013 for characteristics including race (share Black, Hispanic, white), age (young = <27, mid=[27,32], and old= >32), gender (share female), and shares of those with military experience and high education (Bachelors or above).

Figure B.2.3: Cohort Composition



Note: Figure displays the cumulative distributions of officer start ages in cohorts for each Exam 2006, 2010 (main sample), and 2013. The figure illustrates that officers cannot begin at the academy after the age of 40 or before the age of 23 prior to Exam 2013 and 21 in Exam 2013 due to a policy change.

Figure B.2.4: CDF of New Officer Start Ages

	Minority Arrests	Non-Black Arrests	Black Low-Level Arrests	Black Serious Arrests
	(1)	(2)	(3)	(4)
Black	-3.56^{***}	-1.08^{***}	-2.83^{***}	0.08
	(0.64)	(0.24)	(0.50)	(0.23)
Hispanic	-1.71^{***}	−0.15	-1.54^{***}	−0.13
	(0.55)	(0.22)	(0.38)	(0.19)
Asian/Native American	-2.48^{***}	−0.27	-1.91^{***}	-0.42^{*}
	(0.85)	(0.39)	(0.63)	(0.24)
Minority, Start Age [27, 32]	-1.81^{***}	−0.17	-1.20^{**}	-0.54^{***}
	(0.70)	(0.21)	(0.51)	(0.16)
Minority, Start Age >32	-1.87^{***}	-0.34^{*}	-1.36^{***}	−0.28
	(0.49)	(0.20)	(0.39)	(0.17)
White, Start Age [27, 32]	-1.53^{***}	−0.24	-1.21^{***}	−0.11
	(0.55)	(0.21)	(0.42)	(0.12)
White, Start Age >32	-3.75^{***}	−0.47	-2.91^{***}	-0.45^{**}
	(0.69)	(0.30)	(0.49)	(0.18)
Minority x Female	-2.98^{***}	-0.49^{***}	-1.90^{***}	-0.58^{***}
	(0.48)	(0.15)	(0.35)	(0.16)
White x Female	-3.08^{***}	−0.43	-2.17^{***}	-0.66^{***}
	(0.57)	(0.29)	(0.44)	(0.18)
Military	−1.60	−0.48	−1.54	0.36
	(1.08)	(0.39)	(1.02)	(0.41)
High Edu	0.44	0.07	0.44	−0.02
	(0.43)	(0.14)	(0.34)	(0.12)
Exam 2010	-7.34^{***}	-1.88^{***}	-9.19^{***}	3.33^{***}
	(0.93)	(0.22)	(1.03)	(0.32)
Exam 2013	-15.32^{***}	-3.76^{***}	-18.62^{***}	6.22^{***}
	(1.01)	(0.24)	(1.16)	(0.35)
Intercept	-101.24^{***}	-79.62^{***}	-65.59^{***}	24.91^{***}
	(1.31)	(0.47)	(1.31)	(0.52)
R^2	0.32	0.16	0.51	0.47
Num. obs.	2336	2336	2336	2336

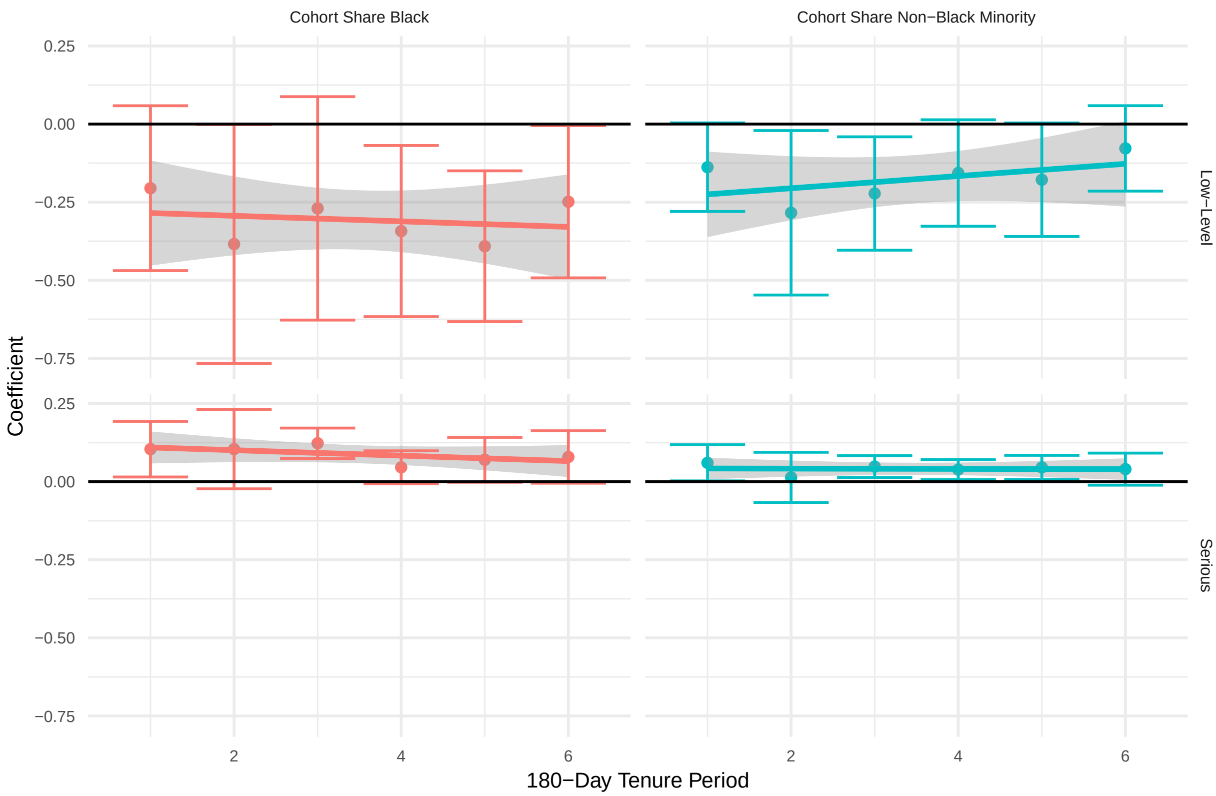
Note: Table displays the linear regression estimates of full sample officer characteristics on their individual propensities to make arrests of various types, recovered using equation 2. High edu corresponds to having a Bachelors degree or above. Non-Black includes white, Hispanic, Asian, Native American, etc.. Coefficients are scaled up to be per 100 shifts for easy of interpretation. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.1: Association between Officer Characteristics and Arrest Propensity (per 100 shifts)

	All	Minority	Black	Non-Black	Non-Black Minority	White
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.34^{**}	-0.31^{**}	-0.26^{**}	-0.09^{*}	-0.06	-0.03^{*}
	(0.16)	(0.15)	(0.12)	(0.05)	(0.04)	(0.02)
Cohort Share Non-Black Minority	-0.21^{*}	-0.19^{**}	-0.15^{**}	-0.06	-0.04	-0.02
	(0.11)	(0.10)	(0.07)	(0.04)	(0.03)	(0.01)
Black	-0.03^{***}	-0.03^{***}	-0.02^{***}	-0.01^{***}	-0.01^{***}	-0.00^{**}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Hispanic	-0.02^{***}	-0.02^{***}	-0.02^{***}	-0.00	-0.00	-0.00^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Asian/Native American	-0.02	-0.02	-0.02^{**}	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Male	0.02^{***}	0.02^{***}	0.02^{***}	0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Start Age	-0.00^{***}	-0.00^{***}	-0.00^{***}	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cohort Size	0.00	0.00	0.00	0.00	0.00	0.00^{*}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	-1.59^{***}	-1.34^{***}	-0.66^{***}	-0.93^{***}	-0.68^{***}	-0.25^{***}
	(0.07)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)
R^2	0.08	0.08	0.08	0.03	0.02	0.04
Num. obs.	940	940	940	940	940	940

Note: Table displays the effect of cohort composition on main sample officer outcomes. The outcomes are individual officer fixed effects, estimated using equation (2), for all arrests, minority, Black, non-Black, non-Black minority (mainly Hispanic), and white; effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.2: Effect of Cohort Diversity on Low-Level Arrest Propensity Across Arrestee Groups - Main Sample



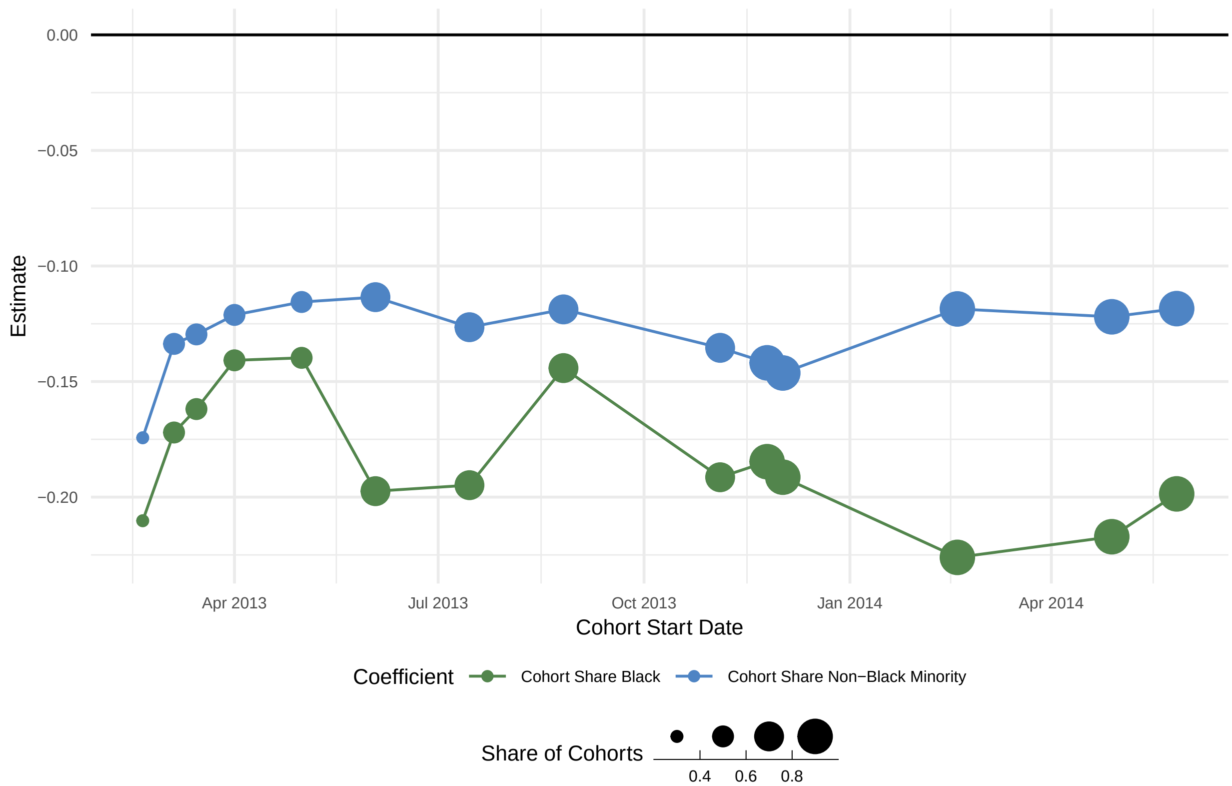
Note: Figure visualized the effects of cohort shares Black, Hispanic, and other non−whites on main sample officer fixed effects for arresting Blacks for serious and low−level crimes over their careers. Unshrunken (raw) fixed effects were recovered by modifying equation 2 such that officers had a separate fixed effect computed for each 180 day period of their tenure. Coefficients are estimated using equation 3 and the main specification (see Columns (9) and (10) in Table 3) separately for each 180−day period fixed effect. Cohort shares are computed as the leave−out mean of the officer's cohort's initial composition. Grey error bars indicate 95 percent confidence intervals based on standard errors clustered at cohort level.

Figure B.2.5: Dynamic Effect of Peer Composition on Propensity to Arrest Blacks

	Shrunken Low-Level Black Arrest Propensity					
	Main Sample			Full Sample		
	Not Guilty	Dropped	Missing	Not Guilty	Dropped	Missing
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	−0.049	-0.044^{***}	−0.011			
	(0.031)	(0.014)	(0.013)			
Cohort Share Non-Black Minority	-0.033^{*}	-0.029^{**}	−0.006			
	(0.017)	(0.012)	(0.006)			
Cohort Share Minority, Start Age <27				−0.001	−0.015	0.009
				(0.014)	(0.010)	(0.007)
Cohort Share Minority, Start Age [27, 32]				−0.021	-0.029^{**}	−0.002
				(0.019)	(0.012)	(0.010)
Cohort Share Minority, Start Age >32				-0.047^{***}	-0.048^{***}	0.009
				(0.012)	(0.009)	(0.008)
Controls	Full	Full	Full	Full	Full	Full
R^2	0.061	0.057	0.032	0.443	0.497	0.049
Num. obs.	940	940	940	2336	2336	2336

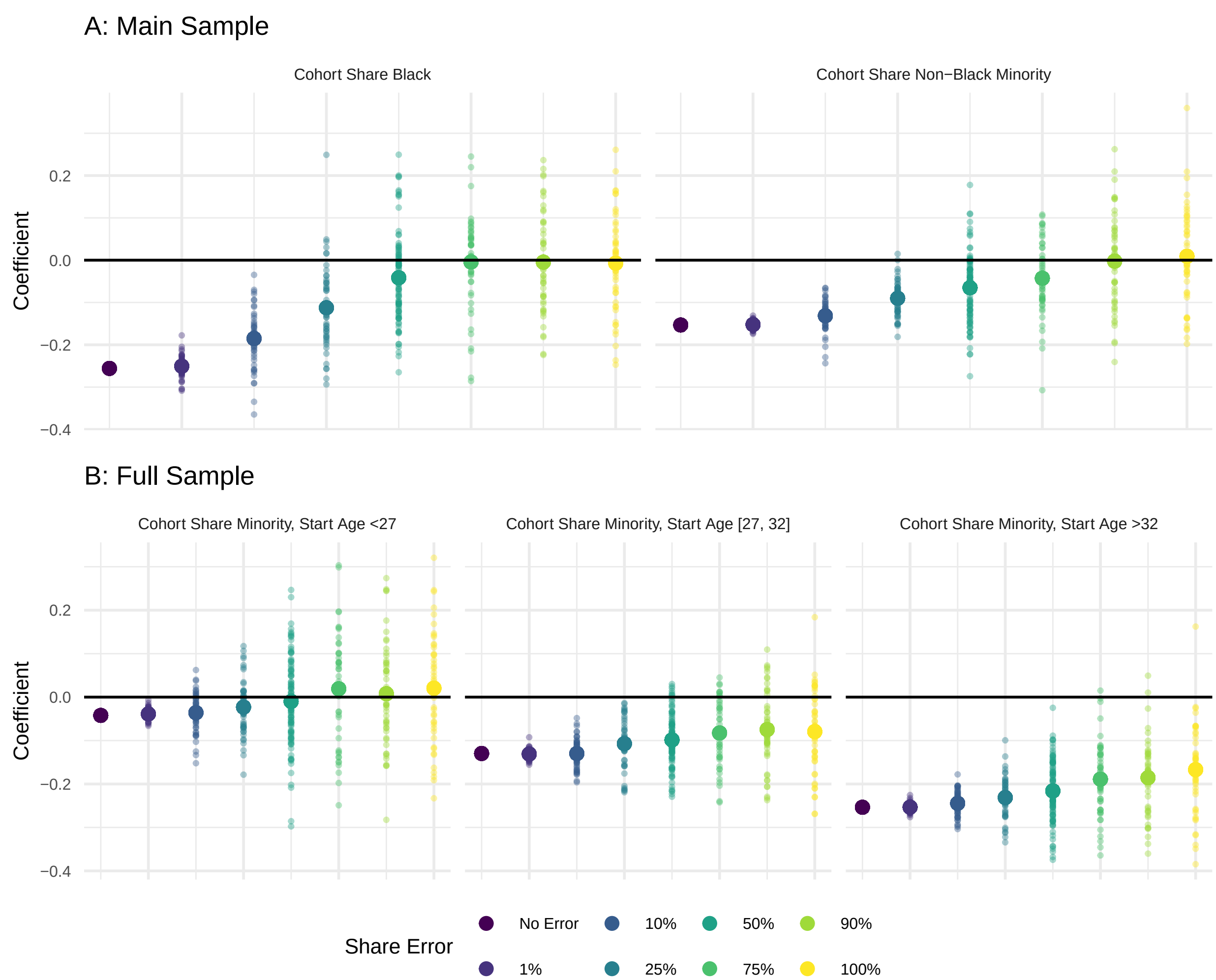
Note: Table displays the effect of cohort composition on main and full sample officers' propensities to low (non-guilty) quality arrest of Blacks for low-level crimes decomposed into not guilty (finding of not guilty), dropped (case dropped/dismissed), and missing (case does not appear in court data). Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.3: Decomposed Effect of Cohort Composition on Low-Quality Arrest Propensity



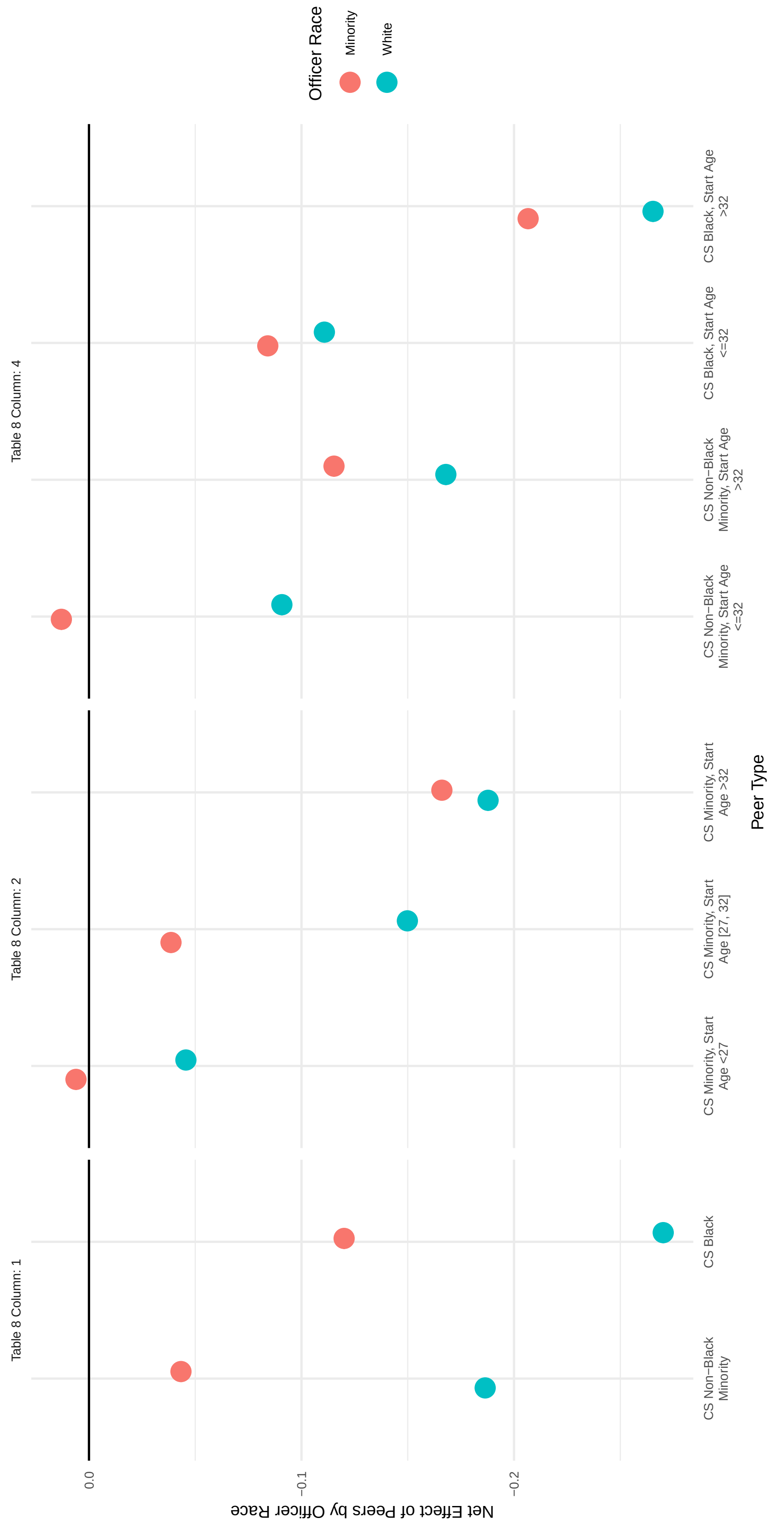
Note: Figure displays coefficients (y−axis) recovered from estimating equation 3 on the main sample of cohorts which started on or before each date (x−axis). The dependent variable is shrunken officer propensity to arrest Blacks for low−level crimes. As the sample size increased (more cohorts are included), the coefficients become more stable.

Figure B.2.6: Coefficient Estimates by Increasing Number of Main Sample Cohorts



Note: Figure visualizes how coefficients change as measurement error is added to peer racial composition. Coefficients are the effects of cohort shares Black and non−Black
sample (panel A) and the effects of cohort share minority by age group for full sample officers (panel B) on shrunken officer fixed effects for arresting Blacks for low−level crim
error is induced by taking the initial sample and assigning racial group (Black, Non−Black minority, and white) based on a uniform random variable for some share ('Share Err
peer compositions are computed. For each share error of observations with measurement error, this exercise is repeated 50 times, and each faint dot corresponds to a particu
(on per value of share error), are the mean coefficients across runs. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as descr
Coefficients estimated using equation (3) using full controls (specification in Column (13) in Table 3 for main sample and specification in Column (5) in Table (4) for the full sar

Figure B.2.7: Change in Coefficients with Measurement Error



Note: Figure visualized the net effects of cohort shares Columns 1,2, and 4 in Table 8 by sample (main and full) for white officers and minority officers. White officer effects are computed by adding the minority officer coefficient to the white officer interaction coefficient

Figure B.2.8: Visualization of Interaction Terms (Table 8)

	Shrunken Low-Level Black Arrest Propensity					
	(1)	(2)	(3)	(4)	(5)	(6)
Cohort Share Black	-0.271^{**}	-0.168^{**}				
	(0.107)	(0.076)				
Cohort Share Non-Black Minority	-0.187^{**}	-0.093				
	(0.079)	(0.061)				
Cohort Share Minority, Start Age >32			-0.134^{***}	-0.127^{**}		
			(0.052)	(0.062)		
Cohort Share White, Start Age >32			-0.057	-0.132^{*}		
			(0.084)	(0.080)		
White x Cohort Share Minority, Start Age >32				-0.008		
				(0.059)		
White x Cohort Share White, Start Age >32				0.134^{**}		
				(0.060)		
Cohort Share Minority and Female					-0.200^{**}	-0.165^{**}
					(0.080)	(0.071)
Cohort Share White and Female					-0.034	0.011
					(0.101)	(0.104)
White x Cohort Share Minority, Female						-0.067
						(0.065)
White x Cohort Share White, Female						-0.081
						(0.093)
Excluded Officers	Minority	Minority	Start Age > 32	Start Age > 32	Female	Female
Sample	Main	Full	Full	Full	Full	Full
Controls	Full	Full	Full	Full	Full	Full
R^2	0.068	0.376	0.400	0.402	0.385	0.385
Num. obs.	482	1161	1756	1756	1841	1841

Note: Table displays the effect of cohort composition on main (Column (1)) and full sample (Columns (2)-(4)) officers' propensities to make arrest of Blacks for low-level crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3). Each column excludes the group of officers which comprise the peer group whose effect is estimated. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3 for main sample with the addition of exam fixed effects for full sample regressions, and it excludes any control which is colinear with the excluded group (e.g., controls for gender are excluded from Column (4)). Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.4: Effect of Cohort Diversity on Arrest Propensity Excluding Peer Officers

	Shrunken Arrest Propensity Black Low-Level			
	Full Sample			
	(1)	(2)	(3)	(4)
Cohort Share Minority, Start Age <27	0.028	0.025	-0.362^{***}	-0.365^{***}
	(0.072)	(0.072)	(0.130)	(0.132)
Cohort Share Minority, Start Age [27, 32]	−0.041	−0.043	-0.421^{**}	-0.424^{**}
	(0.086)	(0.086)	(0.178)	(0.180)
Cohort Share Minority, Start Age >32	-0.150^{***}	-0.151^{***}	-0.510^{***}	-0.510^{***}
	(0.049)	(0.049)	(0.146)	(0.147)
Cohort Share White, Start Age >=27	-0.066^{*}	-0.067^{*}	-0.066^{*}	-0.068^{*}
	(0.039)	(0.039)	(0.038)	(0.038)
Cohort Share Minority and High Edu	−0.085	−0.093		
	(0.058)	(0.068)		
Cohort Share White and High Edu	0.019	−0.031		
	(0.053)	(0.051)		
White x Cohort Share Minority and High Edu		0.011		
		(0.044)		
White x Cohort Share White, High Edu		0.101^{***}		
		(0.037)		
Cohort Share Minority and Military			0.161^{*}	0.179^{*}
			(0.088)	(0.093)
Cohort Share White and Military			-0.190^{***}	-0.211^{***}
			(0.066)	(0.066)
White x Cohort Share Minority, Military				−0.036
				(0.025)
White x Cohort Share White, Military				0.044^{*}
				(0.023)
Controls	Full	Full	Full	Full
R^2	0.430	0.431	0.433	0.434
Num. obs.	2336	2336	2336	2336

Note: Table displays the effect of cohort composition on full sample officers' propensities to make arrest of Blacks for low-level crimes. Arrest propensities are individual officer fixed effects estimated using equation (2) and shrunken as described in Appendix A.6. Effects estimated using equation (3) with additional variables of interest listed in the table. such as addition of terms for cohort shares interacted with an officer being white. High edu refers to having a Bachelors degree or above before the academy. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Full controls refers to the additional controls in Column (13) of Table 3. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.5: Alternative Mechanisms for Effect of Peer Diversity on Arrest Propensity

	Shrunken Black Arrest Propensity							
	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort Share Black	-0.212^{**}	0.036^{***}	-0.120^{*}	0.026^{***}				
	(0.086)	(0.012)	(0.067)	(0.006)				
Cohort Share Non-Black Minority	-0.121^{**}	0.009	−0.043	0.005				
	(0.055)	(0.009)	(0.047)	(0.006)				
Cohort Share Minority, Start Age <27					−0.029	0.006	−0.117	−0.006
					(0.049)	(0.009)	(0.088)	(0.027)
Cohort Share Minority, Start Age [27, 32]					−0.093	0.005	-0.159^{*}	−0.002
					(0.069)	(0.007)	(0.093)	(0.017)
Cohort Share Minority, Start Age >32					-0.151^{***}	0.019^{***}	-0.187^{***}	0.012
					(0.041)	(0.007)	(0.050)	(0.013)
Controls	Full	Full	Full	Full	Full	Full	Full	Full
Sample	Exams 2006, 2010	Exams 2006, 2010	Full	Full	Full	Full	Exams 2006, 2010	Exams 2006, 2010
R^2	0.261	0.123	0.437	0.158	0.440	0.157	0.261	0.121
Num. obs.	1181	1181	2296	2296	2296	2296	1181	1181

Note: Table displays the effect of cohort composition on main and full sample officers' propensities to arrest Blacks for low-level and serious crimes in even and odd columns, respectively, excluding the 2011 cohort which may have mixed 2010 exam-takers and 2006 exam-takers. The propensity is captured by officers' fixed effects using equation (2) and shrunken as described in Appendix A.6. The parameter estimates are based on the specification in equation (3). Full controls refers to the additional controls in Column (13) of Table 3 with exam fixed effects and controls for cohort shares of whites who are start between 27 and 32 and above 32. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.2.6: Effect of Cohort Composition on Arrest Propensity - Full Sample Excluding 2011 Cohort

B.3 Appendix C - Data

The data used in this study were obtained via FOIA request, in collaboration with the Invisible Institute and Chicago Data Collaborative, and generously shared by Rachel Ryley.

Demographics Data on officer demographics were obtained via multiple FOIA request to the Chicago Police Department. These data include information on officers extending as far back as the 1940's to the present (2021). The core demographic data includes name, race (ethnicity), start date, resignation date, and gender. Additional data sets relating to officer's language abilities were obtained for more recent officers (i.e., those in the data for this study), which were used to determine if the officer reported being able to speak Spanish. Similarly, whether or not an officer was in the military was also obtained for the present set of officers. Educational attainment records were also obtained indicating where, when, and what degree (if any) was obtained by each officer– this data is much less complete than other data sets but is most complete for officers starting around the main sample. For simplicity, educational data was summarized for this study as an indicator ("high edu") if the officer had reported obtaining a Bachelors degree or higher (e.g., masters, law degree, doctorate) before they started at the academy.

Salary Salary data, obtained via FOIA to the Department of Human Resources, contains salary, pay grade (rank), and promotion information for officers between 2002 and 2020. This data is important as it allows us to focus on 'regular' police officers, i.e., D1 employees, and filter out promoted employees (sergeants, detectives, etc.). Importantly, this data contains officers' age at hire, allowing for very close approximation of their actual birth date and thus their exact age upon starting at the academy.

Unit History Officers' official unit assignments were obtained via FOIA to the CPD. This data indicates the dates on which an officer began and ended their official assignment to a specific unit.

Daily Assignments On a day to day basis, officers work specific beat assignments (alphanumeric codes that relate to function and location), are on specific watches, are or are not present for duty, are absent for some reason, are assigned to specific cars, and work between specific times.

This information is contained within the daily assignment data, referred to in the text often as "AA" data. This data was obtained for the 22 (25 pre-2013) geographic units focused on in this study via FOIA request (for years 2010-2011 and 2016-2018) and shared by Rachel Ryley (for 2012-2015). Additional information on officer 'roles' were obtained via FOIA request to the CPD which gave descriptions of almost all beat assignment code to clarify their meaning.

Trainings A training data set, supplementary data set to the AA data, was obtained via FOIA request covering the period of the study. Specifically, this contains the name and start time of classes/trainings officers attended. This is particularly useful for identifying which officers were consistently trained together during the academy within their cohorts.

Arrests Data on adult arrests in Chicago were obtained via FOIA request to the CPD. This data includes arrestee information (race, age, gender), identifying officer information, arrest date and time, crime type and description, and the officer's arrest role (primary, secondary, or assisting). The arrest severity (Serious or Low-Level) is by crime type. Serious crimes include all violent and property index crimes, non-index property, and non-index violent crime (such as domestic violence and all forms of sexual assault). See *Crime* for crime code information. All other crimes (e.g., traffic, gambling, prostitution, drug) are considered low-level.

Court Court data from the Circuit Court of Cook County was obtained through collaboration with the Invisible Institute and Chicago Data Collaborative. This data is used to link specific arrests to cases and thus court outcomes (i.e., guilty finding, dropped case, etc.). It contains cases through 2019.

Population Information on district populations is obtained from the 2010 US Census. Median income is based on 2014 ACS estimates.

Crime Raw crime data is obtained from the Chicago Data Portal, downloaded in August of 2020. Crime is classified based on FBI codes into violent, property, and other crime. Violence-related crime FBI codes are 1A/B (homicide/manslaughter), 2 (criminal sexual assault / rape), 3 (robbery), 4A/B (aggravated assault/battery), 8A/B (simple assault/battery). Property-related crime FBI codes are 5 (burglary), 6 (theft), 7 (motor vehicle theft), 9 (arson), 10-13 (deceptive practices/fraud/stolen property), 14 (criminal damage). Index crime codes are 1A, 2, 3, 4A, 4B, 5, 6, 7, 9. All other crimes are classified as other and non-index, e.g., prostitution, gambling, trespassing, narcotics. Arrest data have the same classifications using FBI codes.

Appendix C: Chapter 3

C.1 Data

Complaints Each complaint corresponds to a unique complaint register number, which represents a single incident involving any number of officers. Complainants can initiate their complaint by phone, in person at the oversight agency's location, by mail, with any CPD supervisor at any district station, or over the internet. Moreover, officers can file a complaint against other officers (See [255] for more details).

Serious complaints are composed of allegations against an identified officer involving improper use of force, search, arrest or lock up, or verbal abuse. Allegations in this category tend to be filed by an individual suspected by the police of criminal activity. FPS complaints are those filed against an identified officer that involve a failure to provide service or conduct unbecoming of an officer and are not classified as Serious. FPS allegations tend to be filed by potential victims of a crime or people seeking help from the police.[1] "Other" complaints are allegations with a non-missing finding and an identified officer where the complaint does not fall into either the Serious or FPS category.[2] For "unknown" complaints, all allegations associated with the complaint either have missing findings or no identified officers. Naturally, we will focus our discussion on Serious, FPS, and Other complaints.

Tribune Articles We include frequencies of topics and phrases used in news articles between 2007 and 2018 from the Chicago Tribune, the city's largest newspaper, relating to the Chicago Police Department.The initial sample of Chicago Tribune articles was the first 8,000 articles avail-

[1] [255] shows civilian complainants have different incentives depending on the category of alleged misconduct.

[2] "Other" categories of misconduct may include misfiling reports, owing money to the city, operational personnel violations, or workplace complaints.

able after performing a search on chicagotribune.com/search/ for "Chicago Police Department". Panels B and C in Figure 3.1 present the frequency distribution of articles mentioning the DOJ and the Chicago police union, and the frequency distribution of articles related to gun violence, police accountability, and police misconduct.

C.2 UCR Data

We focus on the Chicago Police Department and local police departments from cities with populations exceeding 250,000 at any time between 2007 and 2019 in the UCR data.

In order to account for seasonal patterns in crime outcomes, we consider the residualized crime rates by regressing crime rates on city-month dummies and using the residuals as the dependent variable.

As noted in [309, 232, 234, 235], the UCR data contains reporting errors that vary across agency. Hence, the data requires thorough cleaning before use.

We delete agencies that are flagged by our procedure as outliers. Following [234] and [235], we use a smoothed version of the measure for the city population reported in the UCR. We exclude agencies that report negative crimes and/or police departments that did not report any crimes in a given year. We only consider local police departments for our analysis (excluding sheriff's departments, U.S. Marshals, state police, and other special jurisdictions). Agencies with missing population data are also dropped.

The resulting data set contains a panel of the Chicago Police Department 65 additional local police agencies from January 2007 to December 2019, depending on the restrictions. Following [231], our main specification only includes large jurisdictions (population $\geq$ 250,000).

C.2.1 UCR Data Cleaning

The annual city population reported in the FBI files tends to jump discretely around census years ([234, 235]). We follow [235] and replace the reported population with a smoothed version. For each city, we fit the population time series using local linear regression and replace the reported population with the fitted values.

Similar to [235], we developed a regression-based procedure to detect cities with extreme outliers and record errors. For each agency, we fit the time series of police reported crimes for assault, burglary, motor vehicle theft, larceny, and robbery using a local linear regression with bandwidth two. We do not use this procedure for murder because it is a relatively rare event. For each crime k, let $y_{k,ct}$ and $y^{pr}_{k,ct}$ be the actual reported crime and predicted reported crime from the local linear regression for agency c during month t. We then compute the following quantity for each crime

$$\delta_{c,k} = max_t \left(\frac{|y_{k,ct} - y^{pr}_{k,ct}|}{0.5 \times (y_{k,ct} + y^{pr}_{k,ct})} \right) \tag{C.1}$$

For each agency we compute

$$\delta_c = max(\delta_{c,assault}; \delta_{c,burglary}; \delta_{c,larceny}; \delta_{c,robbery}; \delta_{c,motor\ vehicle\ theft}) \tag{C.2}$$

We consider an agency to be an extreme outlier or to have a recorded error if $\delta_c > 1.5$. Agencies flagged as outliers are deleted from our data.

C.3 Bounding Treatment Effects using [227]

Setup Our main goal is to understand how the three events of interest which potentially induced variation in misconduct allegations impacted crime, officers' arrests, and use of force in Chicago. For exposition reason, let's assume there is only one event of interest. Using the potential outcomes framework for causal inference ([310, 311, 312]), let $Y_{jt}(0)$ and $Y_{jt}(1)$ be the potential outcomes for agency j at time t where the treatment status, $D_{jt} = 1$ if j experience the event of interest at date t and $D_{jt} = 0$ otherwise. Hence, Potential outcomes and the value of the treatment determine the observed outcome such that[3]

$$Y_{jt} = D_{jt} \cdot Y_{jt}(1) + (1 - D_{jt}) \cdot Y_{jt}(0) \tag{C.3}$$

[3] for exposition, we omit covariates.

For conciseness, we often write $Y_{jt} = Y_{jt}(D_{jt})$. Our goal is to recover the average treatment effect at date t given by

$$ATE_{t,j} = E(Y_{jt}(1)) - E(Y_{jt}(0)) \tag{C.4}$$

Unfortunately, we do not have data for other agencies for reasons we explained in the main text. As an alternative, we use the two previous years as control groups for Chicago during the event year. For example, for the first event occurring in July of 2009, we will use the months around July of 2007 and July of 2008 as control groups. This approach allows us to use a difference-in-difference estimator to recover $ATE_{t,j}$. We estimate the model of the following form

$$Y_{jt} = \alpha_j + \gamma_t + D_{jt}\beta + \epsilon_{jt} \tag{C.5}$$

where Y_{jt} and D_{jt} are the observed outcomes and treament dummies (defined previously); α_j and γ_t are the dummies for each group and time fixed-effects. The error term ϵ_{jt} is assumed to be mean zero conditional on treament and fixed effects. The parameter of interest,β, is the treament effect and is assumed to be constant. The DID estimator is correct under the assumption that in the absence of the event varying le level of oversight, treated and control groups would have experienced the same change in outcomes after the month of the event.

To recover the bounds of the DID, we make the following assumptions from [227] in their empirical application:

1. Bounded intergroup variation:$|Y_{j,t}(D) - Y_{l,t}(D)| \leq \delta_{(jl)t}$

2. Bounded DID variation: $|[Y_{j,t}(D) - Y_{j,e}(D)] - [Y_{j,t}(D) - Y_{l,t}(D)]| \leq \delta_{(jl)(te)}$

such that $D = 0, 1$ and (j, l) are specified groups, (t, e) are specified month, and $(\delta_{j(te)}, \delta_{(jl)t}, \delta_{(jl)(te)})$ are specified positive constants.

For the selecting the bounded parameters, we make the following assumption: (1) $\delta_{(j,l)(te)} = Q_{0.75}(DID))$ to be consistent with the data in each period prior to the event; and (2) $\delta_{(l,j)(t)}$ is such that the maximum of the minimum parameter value of the inter-group variation so that the models are consistent with the data in each month prior to the event. Hence, following [227], we use the following assumptions to construct the bounds of the treatment effects:

1. $|[Y_{j,t}(0) - Y_{j,e}(0)] - [Y_{l,t}(0) - Y_{l,e}(0)]| \leq \delta_{(j,l)(te)}$ where $\delta_{(j,l)(te)} = Q_{0.75}(DID)$ for $t \in [1, ..., 5]$ and $e = -1$

2. $0 \leq Y_{l,t}(0) - Y_{j,t}(0) \leq \delta_{(l,j)(t)}$ for $t \in [1, ..., 5]$

3. $Y_{j,t}(0) \geq Y_{l,-1}(0)$ for $t \in [1, 2]$ and $Y_{j,t}(0) \leq Y_{l,-1}(0)$ for $t \in [4, 5]$