Are "Bad" Cops Better Police?

The Trade-off Between Officer Aggression and Public Safety

Roman Rivera*

February 5, 2025

Abstract

This paper explores the trade-off between police officer behaviors and crime, focusing on the relationship between productive law enforcement, crime reduction, and high-cost behaviors (misconduct and use of force). I use a decade of daily shift data from the Chicago Police Department to estimate the causal effects of individual officers on stops, arrests, use of force, misconduct, and crime. Officer heterogeneity generates significant variation in each outcome. However, officers more prone to costly behaviors are only modestly more productive, as measured by enforcement, and are not better at reducing crime. Furthermore, officer-level activities, such as stops and low-level arrests, have minor direct effects on crime. Finally, I apply social cost estimates to policing activities and crime outcomes. Officers more prone to misconduct and use of force do not prevent enough crime to justify their high-cost behaviors, despite making more arrests and stops. These results challenge the belief that aggressive policing improves public safety.

*For feedback, comments, and advice, I thank Bocar Ba, Patrick Kline, Dorothy Kronick, Raffaele Saggio, Amani Abou Harb, Hadar Avivi, Matthew Gudgeon, Jon Moreno-Medina, Emma Harrington, Brian Bell, Sam Norris, David Card, Steven Raphael, Conrad Miller, Steve Mello, Taeho Kim, Sandra Black, and the participants at the UC Berkeley Labor Seminar, UC Berkeley Goldman Research Seminar, University of British Columbia Economics Seminar, the Texas Economics of Crime Workshop, Princeton IR Section's Northeast Labor Symposium for Early Career Economists, and the Conference on the Economics of Crime and Justice. I thank the Institute for Research on Labor and Employment, UC Berkeley Center for Labor Economics, and Princeton University Industrial Relations Section for their support. This work received support from the Institute for Humane Studies under grant number IHS017836. IRB approval for this study was received from Georgetown University (#STUDY00007559). IRS Postdoctoral Research Associate, Princeton University; Email: romgariv@gmail.com

1 Introduction

Optimal law enforcement balances the beneficial reduction in crime from policing with the direct and indirect costs of policing (Becker (1968)). While the financial costs and social savings of hiring additional officers to reduce crime are well-studied (Chalfin & McCrary (2018), Mello (2019)), a growing body of work documents the extensive costs associated with policing activities. These include the costs of settlements from misconduct (Ouss & Rappaport (2020)), lost educational attainment from use of force (Ang (2021)), and downstream consequences of arrests (Mueller-Smith (2015), Dobbie et al. (2018), Agan et al. (2023)). A growing body of work studies how to reduce these costly activities through recruitment, departmental policy, and training (e.g., Owens et al. (2018), Dube et al. (2023)). Nevertheless, public and academic discourse often assume that proactive or aggressive policing produces more effective law enforcement (Corman & Mocan (2005), Prendergast (2007), Shi (2009), Mac Donald (2016), Linnemann (2022), Denman (2023)). This implies a potential trade-off, as reducing the costly behaviors of officers may also reduce effective policing and increase crime, which motivates the need for understanding the direct links between officer behavior, policing outcomes, and crime.

Research on policing and crime has long explored the relationship between crime and enforcement, often in the form of the quantity of police and police presence in specific locations (Draca et al. (2011), Chalfin & McCrary (2017)). Proponents of more aggressive policing argue that strict enforcement against low-level offenses reduces crime (Kelling & Wilson (1982)), and existing work finds that more police reduce crime at the expense of more low-level arrests (Chalfin et al. (2022)). Prior attempts to study the relationship between crime and costly behaviors (e.g., misconduct or low-level arrests) have used large-scale changes in the forms of oversight, scandals, or reforms (Prendergast (2001), Shi (2009)) when many factors change at once (Rivera & Ba (2023)).[1] A separate literature on officer-level

[1] See also Devi & Fryer (2020), Premkumar (2022), Mikdash & Zaiour (2023), and Campbell (2024), Tebes & Fagan (2025), and Parker et al. (2024).

heterogeneity shows that policing outcomes are strongly influenced by which officer is working, including arrests (Weisburst (2022)), misconduct (Rozema & Schanzenbach (2019)), searches (Feigenberg & Miller (2022)), ticketing (Goncalves & Mello (2021)), and use of force (Adger et al. (2022)). However, the direct link between crime reduction and officer-level activities is missing. As a result, it remains unclear whether there is an actual relationship between an officer's effectiveness at reducing crime and their propensity to engage in more aggressive and socially costly policing activities.

This paper bridges these gaps by documenting the relationship between officer effectiveness in reducing crime and their propensity to engage in enforcement activities and costly behaviors, such as misconduct and use of force. Using extensive shift-level data on Chicago Police Department officers from 2010 to 2019, I estimate the effects of individual officers on policing activities, including stops, arrests, use of force, and misconduct complaints. I extend the canonical worker-firm design (Abowd et al. (1999)) and leverage the CPD's fixed rotation schedule and officers moving between assignments to disentangle officer-specific effects from the confounding influence of assignments. Then, I estimate bias-corrected variance and cross-outcome covariances in officers' activity effects (Kline et al. (2020), Lachowska et al. (2023)). I also estimate individual officer effects on crime rates and the degree to which individual officers matter for crime reduction by exploiting variation in which officers work on given days and times due to the rotations calendar and shift start/end time information. Next, I connect crime with policing activities. I estimate the relationship between officer-activity effects and crime rates (e.g., the effect of higher use-of-force officers on violent crime) and between policing activities and crime (e.g., the elasticity of property crime with respect to street stops). Finally, I apply incidence costs to crimes and to policing activities to determine if more costly officers produce social savings through greater crime reduction.

I find that officer heterogeneity is economically important for stops, arrests, uses of force, and complaints for misconduct. One standard deviation (SD) in officer fixed effects is generally close to or exceeds 75% of the mean for each activity. Notably, officer effects

explain more variation than do assignment effects for all outcomes other than index arrests, which are more serious and less discretionary. This is particularly true for uses of force and complaints, where variation across officer effects is generally twice as large as variation across assignment effects, suggesting which officer is working is a much larger factor than their role and location in determining the most costly outcomes.

Despite officer heterogeneity influencing each policing activity, I find that officer effects for different activities tend to be only modestly correlated. Officer effects for complaints and use of force have the highest cross-activity correlation, ranging between 0.54 and 0.64. However, complaint effects are much more modestly correlated with low-level arrests ($0.28 - 0.32$) and weakly correlated with street and traffic stops and index arrests (all < 0.2). Use of force effects are modestly correlated with effects for index arrest ($0.32 - 0.39$) and nonindex arrests ($0.36 - 0.42$) but more weakly correlated with street and traffic stops (all < 0.3). These results are inconsistent with a model in which officers who are the most productive and exert higher effort, resulting in more enforcement, are also those who engage in the most costly behavior.

I find that officers also differ in terms of their effects on crime. Holding fixed officer quantity, variation in team composition (the exact officers working in a given district-hour) explains modest but economically significant shares of variation in crime. Team composition has a comparable influence in magnitude to the effect of officer quantity (team size) from Chalfin & McCrary (2018) and Mello (2019) for property crime, though less than half the size of quantity-effects on violent crime. Furthermore, officers also influence the frequency of 911 calls, suggesting these effects measure underlying influences on crime, with additional results consistent with differential effectiveness in deterrence.

I find a weak relationship between officer-activity effects and their effects on violent and property crime, and I find no evidence that officers with higher complaint effects or force effects are significantly better at reducing violent or property crime. I also estimate the elasticities of violent and property crime with respect to stops and arrests. These activities

have no statistically significant effect on violent crime, with all estimates smaller than -0.07 and confidence intervals rejecting elasticities less than -0.16. For property crime, I estimate elasticities for nonindex arrests and street stops of -0.1 and -0.053 (both $p < 0.01$), rejecting effects less than -0.15 and -0.09, respectively. For index arrests and traffic stops, I find relatively precise null effects. Even the lower bounds of the elasticity estimates are significantly smaller than the estimates of elasticities of officer quantity on crime (Weisburst (2019), Mello (2019)). Additionally, these estimates complement those of Cho et al. (2023), which infers elasticities using arrest declines following officer deaths, though I directly estimate more precise elasticities for a larger range of policing activities.

Finally, I apply cost estimates to both crimes and officer activities and compute how total costs change if higher use of force and complaint officers are replaced with average officers, resulting in lower force and misconduct but also less enforcement due to the positive cross-activity correlations. In the preferred specification, reducing an officer's complaint or use of force effect by 1 SD generates at least $9,000 in savings over 100 shifts after accounting for changes in correlated enforcement behavior. This is for two reasons. First, 'bad' cops are not significantly more active police because enforcement effects are not strongly correlated with misconduct and force effects. Second, policing activities such as stops and arrests, traditionally viewed as clear metrics of productivity, effort, and crime-fighting, do not actually have large direct effects on crime.

In sum, while officer heterogeneity is significant for officer-level outcomes and even crime, the results suggest only a small trade-off between officers' tendency to engage in costly behaviors and their propensity for policing outcomes traditionally viewed as productive (i.e., arrests and stops). Nevertheless, officers' ability to reduce crime is largely unrelated to their activity metrics, and I can rule out even relatively small elasticities of crime with respect to low-level arrests and stops. These results suggest that reducing costly policing behaviors through training or recruitment is unlikely to produce out-sized social costs through increased crime or fewer important arrests. More broadly, the features that make a 'good' officer are

not aligned with popular belief or most existing theories of policing, suggesting significant ground for future research.

This paper contributes to the broad literature in labor economics on the importance of heterogeneity across firms, teachers, and bureaucrats on a range of economic and educational outcomes (Card et al. (2013), Chetty et al. (2014), Rose et al. (2022), Best et al. (2023)), by applying the canonical Abowd et al. (1999) framework to policing, a environment in which worker discretion plays a critical role.[2] With respect to police, existing work shows the importance of officer-level heterogeneity on specific dimensions of policing using 'shrunken' estimates of officer effects (Rozema & Schanzenbach (2019), Adger et al. (2022), Weisburst (2022)). I extend this literature by providing consistent estimates of officer heterogeneity across the commonly studied policing activities and estimating the relationship between these dimensions of officer behavior.[3] Additionally, I contribute to both the officer heterogeneity literature and the large literature on the effect of policing quantity on crime (Chalfin & McCrary (2017)), by estimating the role of officer heterogeneity on crime, showing that, beyond quantity, officer 'quality' is important in reducing crime while also showing that the elasticities of crime with respect to traditional metrics of police performance are relatively small (Mas (2006)).[4] This complements the literature in labor economics on team production (Jaravel et al. (2018), Chan (2021), Chen (2021)) by extending the Bonhomme (2021) framework to connect teams of officers to local crime rates.

The paper proceeds as follows. Section 2 discusses background and data. Section 3 introduces the conceptual framework. Section 4 introduces the empirical strategies. Sections 5 and 6 present (co)variance estimates for policing activity and crime effects. Section 7 connects officer activity and crime. Section 8 discusses social costs. Section 9 concludes.

[2]See also Song et al. (2019), Frederiksen et al. (2020), Bonhomme et al. (2022), Chetty et al. (2014), and Jackson (2018).

[3]The variance of shrunken officer effects (using Empirical Bayes shrinkage factors) will under-estimate the true variation in officer effects (Walters (2024)), while the Kline et al. (2020) bias-correction provides unbiased estimates for the true variance of officer effects.

[4]See also, Levitt (2002), Di Tella & Schargrodsky (2004), Klick & Tabarrok (2005), Evans & Owens (2007), Draca et al. (2011), Donohue et al. (2013), Weisburst (2019),Mastrobuoni (2019), Chalfin & McCrary (2018), Cheng & Long (2018), Mello (2019), Cox & Cunningham (2021), Chalfin et al. (2022).

2 Setting and Data

2.1 Background on the Chicago Police Department

The Chicago Police Department (CPD) is the 2nd largest police department in the U.S., and it polices a large and diverse population of around 2.7 million people. There are over 10,000 sworn officers in the CPD at a given time. For an individual officer, much of where, when, and what day they will work is predetermined. Within a given year, their unit (where and in what role they work), watch (what time they work), and day off group (which day they work) were already determined at the end of the prior year. Changes to an officer's stable unit/watch/day off group occur through a bidding process based on tenure (Ba, Bayer, et al. (2021)). A strong union protects CPD officers against violations of their contracts.

The main organizational structure within the CPD is the 'unit', and officers are assigned a stable unit to work in for a given year. Each unit performs different functions, ranging from general policing of a specific area ('patrol units') to detective units, SWAT units, and training units. The most common units are the 'patrol' or 'geographic' units which map to a specific police district — a large area of the city (around 10 square miles) similar to precincts in other cities. There are currently 22 districts, and officers assigned to a geographic unit are generally responsible for patrolling, answering 911 calls, and serving on special and tactical teams within the corresponding district. Figure A.1 displays a map of CPD districts and neighborhoods from Markovic (2009).[5] Other units can map to more generalized areas encompassing multiple districts (e.g., detective or gang enforcement units) or no specific location (e.g., legal affairs or training). While these non-patrol units are more numerous, they are much smaller and include specialized functions such as the Narcotics Division, Gang Enforcement, and the Traffic Section.

Officers also work a specific 'watch', which generally corresponds to when they work.

[5]During this study, Chicago is divided into 22 (2013-2019) or 25 (2009-2012) police districts. The 2009 district map changed slightly throughout 2012, during which three units, 13, 21, and 23, were absorbed by larger districts, primarily districts 12, 2, and 19, respectively. Districts are subdivided into geographic beats (usually between 15 and 20), each around 1 square mile).

Most officers work watches numbered 1-4 and shifts lasting 9 hours. Watches 1 (starting around 10 pm), 2 (starting around 7 am), and 3 (starting around 2 pm) correspond to the regular shifts. Watch 4 is for tactical team officers, as studied by Kim (2024).

Within a given year, the rotations calendar determines which days an officer works based on their day off group (DOG). DOGs determine the rotation of officers working on a given day. Most officers work on a 4-2 rotation (4 days on duty followed by 2 days off duty).[6] Importantly, officers cannot easily or systematically manipulate days on duty. Furthermore, because of overlapping DOGs, the groups of officers working will change from day to day.[7]

When on duty, an officer is assigned to a specific 'beat code', an additional level of specificity for their role and/or specific location on that day beyond their unit and watch. For example, an officer may be a 'beat' officer, meaning they patrol a specific geographic beat (around 1 square mile area within a district); or they may be a rapid response officer responding to high-priority 911 calls within a district; or they may be a bike officer, special mission officer, school officer, or assigned to desk duty.[8]

2.2 Data and Summary Statistics

The primary data for this study covers the daily activities of Chicago Police officers over ten years (2010 - 2019), which I split into three sample periods: 2010-2012, 2013-2015, and 2016-2019. There are two levels of analysis involving distinct but similar samples of officers and units: officer-shift and district-hour. Tables 1 and 2 display their respective summary statistics. Both analyses rely on the daily assignment and attendance records for all units in

[6] In 2009 and before, officers typically adhered to a 6-2 rotation schedule, working six consecutive days followed by two days off, but this was switched to shorter rotations (most commonly 4-2) in 2010 (JG0304 (2009)).

[7] For example, officers with DOGs A, B, and C may work on a specific Monday, while officers with DOGs B, C, and D work on Tuesday, and officers with DOGs C, D, and A work on Wednesday. See A.2 for an example of the CPD's operations calendar.

[8] Assigned beats can map directly to geographic beats, but this is not the norm. Instead, an assigned beat denotes a general area and a role. Officers can influence which beat they work within their unit and watch. However, they generally work the same or a small set of beats on a day-to-day basis, and the exact assignment of officers to beats is determined by upper commanders within the unit and watch (e.g., sergeants and lieutenants).

the CPD, initially covering over 15,000 unique officers working over 10 million shifts. This data contains officer-by-shift information, indicating whether the officer was present for duty, their start and end times, rank (e.g., police officer, sergeant, detective), and assigned beat, watch, and unit, among other information. I focus on Police Officers which are the lowest rank in the CPD (i.e., not a detective or sergeant) and the vast majority of sworn CPD members (around 80%).

2.2.1 Officer-Shift Level

At the officer-shift level, a single observation corresponds to one officer working on a given day, watch, unit, and beat. Along with a few other restrictions, the analysis sample contains officer-shift observations in which officers are present for duty in active assignments (e.g., not missing, in training, or desk duty). As shown in Column (1) of Table 1, there are a total of 12,944 unique officers working a total of 11,993,940 shifts in 54 unique units.

The average officer within a sample period is observed for just over 500 shifts (2010 - 2012, 2013 - 2015) and just over 600 shifts (2016 - 2019). I define an 'assignment' as the interaction of unit, watch, and beat type. Almost all officers ($> 97\%$) in each sample period switch between assignments, with around 60% of an officer's shifts are in their most common assignment, with 92% occurring in their most common unit.[9]

To measure the many dimensions of officer-level activity, I collect data on many of the policing outcomes considered in policy and academic research (stops, arrests, use of force, and misconduct) to the assignment data. Panel D of Table 1 summarizes the means of the primary outcomes per 100 shifts. First, stops are classified as either 'street' stops or 'traffic' stops. Both types are a metric for contact with suspects and non-suspects alike, and they are more frequent than other outcomes, averaging 35 per 100 shifts and 13 per 100 shifts for street and traffic stops, respectively. Notably, in 2016, major reforms resulted in a significant

[9]This study focuses on a sample of 54 units in which officers are sufficiently active, in that they engage frequently in policing activity (stops, arrests, etc.), and Figure A.3 displays these units by name and number along with the number of officer-shift (day) observations per unit in the sample.

decline in street stops and a rise in traffic stops (Hausman & Kronick (2023)).

Next, I link data on officers' arrests and divide arrests into two general categories: arrests for index crimes, 'index arrests', (violent: murder, robbery, criminal sexual assault, aggravated assault; property: burglary, motor vehicle theft, larceny/theft, arson) and arrests for non-index crimes, 'nonindex arrests', (e.g., drugs, prostitution, simple battery, vandalism, municipal code violations, traffic offenses). Nonindex arrests are considered more discretionary, often do not have a victim, and are more common at 11 per 100 shifts, whereas all index crimes have a victim, are considered less discretionary, and are less common at around 1.6 per 100 shifts.

I also link data on uses of force by officers, ranging from relatively minor force (e.g., pressure points, defensive techniques) to gunshots and tasers (which are relatively rare). Uses of force occur at around 0.36 per 100 shifts.

I measure misconduct using complaint data, which contains allegations of misconduct filed against police officers by civilians ($> 85\%$) or fellow officers. Complaints declined significantly from 2016 and beyond from around 0.24 down to around 0.1 complaints per 100 shifts.[10]

2.2.2 District-Hour Level

The second analysis focuses on connecting teams of officers with reported crimes. I construct the team of officers working in the same district at the same time (hour) using unit and the start and end time information in the assignment data — e.g., all officers present for duty in district 12 on June 15th, 2013, from 2-3 pm. This sample focuses on officers assigned to patrol/geographic units which directly map to districts (around 80% of officer-shifts). While I only count police officers working on a team (see Appendix E for additional filters), I do record the number of other ranks active by district-hour to use as controls (e.g., the number

[10]Due to changes in the oversight regime following the 2016 reforms, I use complaint data obtained in 2016 from the older regime (IPRA) for years before 2016 and complaint data obtained in 2023 from the new regime (COPA).

of detectives).

As shown in Table 2, across all sample periods there are 1,836,414 teams, equivalent to district-hour observations. There are a total of 11,468 unique officers in the teams data pooled across periods and 25 unique units (the patrol / geographic units), with the average officer being a part of around 3,800 teams (approximately 422 9-hour shifts) in 1.3 unique units within each sample period. The average team has 49 active officers.[11] Additionally, the size of teams can vary significantly within the same district over the course of the day (Figure B.3).

I map the total number of reported crimes reported when a given team is working by summarizing time- and location-stamped crime data to district-hour level, obtained from Chicago's public data portal. Index property crime (e.g., burglary, motor vehicle theft, theft/larceny) is around 4 times as common as index violent crime (e.g., murder, criminal sexual assault, robbery, aggravated assault) at 0.49 per district-hour compared with 0.13 per district hour. Murder is the rarest crime, by almost an order of magnitude, at only 0.0026 per district-hour, while theft (0.339 per district-hour), nonindex simple violent crimes (0.302), and miscellaneous nonindex offenses (0.42) are very common. Drug crimes are also common (0.123), though they are entirely driven by officer-initiated arrests and clear by arrest at nearly 100% unlike other crimes. In contrast, only 9% of property and 17% of violent crimes are cleared (see Table A.1 for clearance rates).

3 Conceptual Framework

This study aims to determine if officers who engage in more aggressive and socially costly policing also achieve more substantial reductions in crime and thus are "better" officers on net. Following the canonical Becker (1968) framework, optimal policing balances policing's

[11]Though the total number of officers is around 54 a few officers are considered inactive (not on the team) as they are in training or assigned to desk duty. Around 75% of the team are regular watch officers (watches 1 - 3), while most of the remaining officers are tactical team officers (watch 4). In years with overtime information (2014 - 2019), I also include overtime officers as team members, though there are only around 4 on the average team in 2016 - 2019.

beneficial reductions in crime with its associated costs, such as officer salaries, overtime, and the negative spillovers of their actions (e.g., settlements for misconduct, lost wages of an arrestee, etc.). This section develops a conceptual framework to evaluate these trade-offs by linking the effect of individual officers on behaviors to their effects on crime.

Consider the population of police officers, indexed by i, assigned to work in assignment J on date and time t. Let $y^k_{i,j,t}$ be the potential outcome for some specific policing activity k (e.g., the number of uses of force, street stops, false arrests, or civilian interactions). The potential outcome can be expressed as:

$$y^k_{i,t,j} = \theta^k_i + \gamma^k_{J(i,t)} + \beta^k x_{i,j,t} + \epsilon^k_{i,j,t}$$

where θ^k_i is officer i's fixed effect or their influence on the activity — determined by their fixed factors such as their race, gender, or experiences in the police academy and training (Ba, Knox, et al. (2021), Hoekstra & Sloan (2022), Adger et al. (2022), Rivera (2022)). Similarly, $\gamma^k_{J(i,t)}$ captures the amount of the activity that occurs due to the assignment $J(i,t)$ regardless of who works (e.g., some assignments involve more traffic stops than others), $\beta^k x_{i,j,t}$ captures time-varying factors (e.g., there is more crime on New Years than, during the summer, or when wages fall, meaning more opportunities for arrests), and $\epsilon^k_{i,j,t}$ reflects the residual variation.

These activities may influence crime rates where and when the officer works through various mechanisms. For example, theories such as Broken Windows Policing (Kelling & Wilson (1982), Corman & Mocan (2005)) suppose that low-level enforcement activities reduce crime. Let $y^c_{d,h}$ be the potential outcome for the amount of crime of type c (e.g., robbery, theft) occurring in location d at time h:

$$y^c_{d,h} = \Omega^c_{\tau(d,h)} + \beta^c n_{d,h} + \gamma^c_{d,h} + \epsilon^c_{d,h}$$

where $\Omega^c_{\tau(d,h)}$ is the effect of the team of officers τ working in location d at time h — meaning

that on day-time t officer i's assignment $J(i,t)$ is within location d and time h. $\beta^c n_{d,h}$ captures the effect of the quantity of officers working on the team of officers ($n_{d,h} = |\tau_{d,h}|$), $\gamma^c_{d,h}$ captures location- and time-specific factors (e.g., socioeconomic conditions or seasonal crime patterns), and $\epsilon^c_{d,h}$ is residual variation from unobserved factors.

The influence of team composition, $\Omega^c_{\tau(d,h)}$ or 'who' is working — holding fixed the well-documented effect of officer quantity on crime (Chalfin & McCrary (2017)) — depends on the individual contributions of team members to their policing activities, or $\theta^k_i \forall i \in \tau_{d,h}, k$. For example, some teams may be better or worse at deterring crime because their members are more aggressively stopping would-be criminals.

Despite not being directly observed, the influence of an individual officer on the crime rates where they work (officer "quality"), is evidenced by anecdotes by police (Moskos (2008)).[12] While $\Omega^c_{\tau(d,h)}$ may be a highly non-linear function of officer effects across a wide range of activities, let ω^c_i be officer i's average causal effect on crime of type c from adding them to a team and excluding the quantity effect of an additional officer: $\omega^c_i = \mathbb{E}[y^c_{\tau(d,h)\cup i} - y^c_{\tau(d,h)}] - \beta^c$.

If $Var(\omega^c_i) \approx 0$, then individual officer quality is relatively uniform, suggesting that the aggregate effects of policing on crime are driven primarily by officer quantity rather than the effects of differential behaviors across officers. Conversely, if variation across ω^c_i is sufficiently large, then this implies impactful heterogeneity in officer quality.

Determining whether aggressive policing is justified requires evaluating the trade-off between the costs of policing activities and the benefits of crime reduction. Let $w^k \geq 0$ and $w^c \geq 0$ denote the social costs for activity k and crime c, respectively. An officer is socially beneficial if their effects on crime costs (e.g., a prevented murder is more valuable than a prevented jay-walking) outweigh the social costs associated with their activities:

$$\sum_k w^k \times \theta^k_i < -1 \times \sum_c w^c \times \omega^c_i$$

[12] In a letter to his superior, Baltimore police officer Moskos states: "I believe the effectiveness [sic] of my work can best be judged by the low level of crime on my posts during my working hours" (Moskos (2008), page 156).

This condition compares the weighted social costs of officer-activity effects to their weighted benefits of crime-reduction effects. For a single dimension of officer activity, this trade-off may be obvious. For example, an additional street stop may incur costs through lost time and stress but may also deter enough crime to make it net cost-saving, while misconduct may have no direct crime reduction effect but impose significant social costs. However, these dimensions of officer behaviors may be interrelated. For example, officers who make more stops may also receive more complaints due to higher effort. Understanding the relationship between different dimensions of officer activities (e.g., $Cov(\theta^{\text{Stop}}, \theta^{\text{Misconduct}})$) and the relationship between officer activity and crime effects (e.g., $Cov(\theta^{\text{Stop}}, \omega^{\text{Robbery}})$) is vital for determining whether tolerating costly officers is justified by their contributions to public safety.

4 Empirical Strategies

The empirical analysis aims to measure officer effects on each dimension of their activities and effects on crime, then utilize these estimates to recover a number of parameters of interest relating to the relationship between officer behavior and crime reduction.

4.1 Estimating Activity Effects

First, the goal is to estimate the causal effect of each individual officer on each officer-level outcome or 'activity' (stops, arrests, use of force, misconduct). I employ an Abowd et al. (1999) -style (AKM) framework to disentangle the individual officer's influence from other factors, in particular, their assignments. Given officer i working in assignment J on day t, the outcome $Y^k_{i,t}$ (e.g., number of index arrests made by i) for officer i on date-shift t is the sum of officer-specific effect θ^k_i, assignment-specific effect $\psi^k_{J(i,t)}$ (unit-watch-beat type), time-varying observable factors $x_{i,t}$ (date-specific fixed effects), and unobserved factors $\epsilon^k_{i,t}$. I estimate the following equation separately across three sample periods (2010 - 2012, 2013 -

2015, 2016 - 2019) allowing for officer and assignment effects to evolve over time:

$$Y_{i,t}^k = \theta_i^k + \psi_{J(i,t)}^k + \beta^k x_{i,t} + \epsilon_{i,t}^k \tag{1}$$

Identifying variation for the parameters in this model comes from officers working in different assignments over time, producing connections between officers and assignments. Within each period, the analysis sample is the largest connected set of officers and assignments (Abowd et al. (2002)), which is robust to the removal of any single officer (Kline et al. (2020), Bonhomme et al. (2022)), and it contains almost all observations ($\geq 95\%$) as shown in Table B.1.

Following Abowd et al. (1999) and Bonhomme et al. (2022), I assume the following mean independence condition is satisfied:

$$E(\epsilon_{i,t}^k | X, \psi^k, J, \theta^k) = 0 \tag{2}$$

or that conditional on officers, assignments, and time-varying observables, the unobserved influences ('shocks') on outcomes, $\epsilon_{i,t}^k$, are mean independent of past and future assignments. Such independence satisfies the exogenous mobility assumption of Card et al. (2013), allowing officers to sort into assignments based on their fixed characteristics but not systematically choose assignments based on contemporaneous shocks. Two potential violations of the exogenous mobility assumption would be evidenced by officers systematically sorting into assignments based on match-specific values, meaning increases/decreases in outcomes from moves would be asymmetric, and that these match-specific components either 'drift' as officers learn about their assignments over time or that officers are choosing when to switch or when to be on duty (analogous to selecting the 'off duty' assignment) based on unobserved shocks.

I test for these violations, which are discussed more in Appendix B. First, officers cannot easily select days-on due to the operations calendar (Ba, Knox, et al. (2021), Gudgeon et al. (2023)), which would allow for violations of exogenous mobility. I validate the strength

of the rotations calendar by constructing the set of officers who are expected to be working on a given day based on the rotations by taking each officer's first observed spell of 4 days on then 2 days off and projecting that pattern forward throughout the entire year (or, in a minority of cases, 5-2). I then compare this simple predicted rotation with the observed rotation of officers and find that it is accurate despite not ignoring all vacation days, injuries, training, potential for switching units, or changing DOGs. As shown in Figure 1, this simple method accurately predicts 85% of observations early in the year and decreases to 70% by the end of the year. For reference, if officers randomly worked 4 out of 6 days, the share of correct predictions would be around 45%. In later robustness tests, I employ this predicted set of officers on duty to ensure deviations from the schedule are not driving results.

Second, I use the traditional event study plots following Card et al. (2013) to test for asymmetric match-specific components and match-specific components drifting. Figure 2 displays the average outcomes of officers around a switch between assignments with assignments grouped by their average quantile of that outcome (e.g., Q4 of nonindex arrests to Q1 of nonindex arrests). Consistent with symmetric match-specific components, officers moving from higher to lower outcome assignments see similar decreases as the increases of those moving from correspondingly lower to higher outcome assignments. For some outcomes, I do find pre-trends prior to switches, though they are uncommon, parallel across switch types (e.g., switches from Q4 to Q3 to Q2 or Q1 in street stops all display a similar trend), small in magnitude, and driven by the shift right before the switch. To ensure such violations are not driving results, I provide several relevant robustness specifications in Section 5.3.

4.2 Estimating Crime Effects

To estimate the influence of an individual officer on crime where and when they are working (ω_i), I model the number of crimes of type c (e.g., robbery, theft) in district d in day-hour h as an additive function of the influence of the team of officers working $\Omega^c_{\tau(d,h)}$, assignment-time specific effects $\gamma^c_{w(d,h)}$, and additional observables $X_{d,h}$ (e.g., the total number of officers

working in various roles), and unobserved factors $\epsilon^c_{d,h}$.[13] The team's effect on crime and other aggregate outcomes, $\Omega^c_{\tau(d,h)}$, is a function of the individual officers who are working together at that time. While it is possible to model the team production (or prevention) of crime to be a highly non-linear function, I make the simplifying assumption that a team's effect on crime is an additive function of its composite members, similar to Bonhomme (2021):

$$y^c_{d,h} = \underbrace{\frac{1}{N_{\tau(d,h)}} \sum_{i \in \tau(d,h)}^{N_{\tau(d,h)}} \omega^c_i}_{\Omega^c_{\tau(d,h)}} + \beta^c X_{d,h} + \gamma^c_{w(d,h)} + \epsilon^c_{d,h} \tag{3}$$

Note that ω_i is the effect of including i in the team relative to some reference officer, as opposed to the effect of adding officer i to a team after accounting for quantity effects. Similar to the model in equation (1), identifying variation comes from officers moving between district-time fixed effects, and the identifying assumption for equation (3) is that unobserved shocks to crime rates are mean independent of past and future observables, assignments, and teams, or strict exogeneity: $E(\epsilon^c | X, \Omega^c, \gamma^c) = 0$. The ideal district-time controls (interacting unit, year, month, day of week, and hour fixed effects) would isolate variation in team composition based on officers starting and ending their shifts and officers working on a specific date but not another (e.g., the first Tuesday but not the second Tuesday of a month) due to predetermined variation in their rotations calendars. However, this proves infeasible due to the high dimensionality of γ^c relative to the number of observations.

Instead, I adopt a first-differences estimator and identify ω_i's solely through variation in the team composition produced by officers starting and ending their shifts within a shift. This removes the need for estimation of unit-time fixed effects and simplifies estimation at the cost of only using the variation produced in officers' first and last hours of work for identification:

[13] In practice for all specifications, $X_{d,h}$ includes second degree polynomials of the number of officers in the team, and second degree polynomials of other policing counts (number of officers working on tactical teams, working overtime on specific teams, working as desk officers, detectives, sergeants, detectives, etc.) — these officers (e.g., probationary officers, detectives, desk officers, overtime task forces) are excluded from the team composition.

$$\Delta y^c_{d,h} = \Delta \Omega^c_{\tau(d,h)} + \beta^c \Delta X_{d,h} + \Delta \epsilon^c_{d,h} \tag{4}$$

where $\Delta a_{d,h} = a^c_{d,h} - a^c_{d,h-1}$. Additionally, using first differences requires a weaker identification assumption than strict exogeneity: $\mathbb{E}[\Delta \epsilon^c_{d,h} | \Delta \Omega^c_{\tau(d,l)}, \Delta X_{d,l}] = 0$.

I employ the pruning algorithm of Bonhomme (2021) to produce the set of officers for which ω_i can be identified, which includes over 98% of observations (Table B.2).[14]

4.3 Parameters of Interest

There are five sets of parameters of interest utilizing officers' causal effects across outcomes. First, the degree to which differences across individual officers matter for a specific outcome, or 'officer heterogeneity,' is captured by the variance in officer effects. For officer-level outcomes, this corresponds to $Var(\theta^k)$, while, for team-level outcomes, it is the variation in composite team effects $Var(\Omega^k)$. Even if the estimates $\hat{\theta}^k_i$ and $\hat{\omega}^k_i$ are unbiased, $Var(\hat{\theta}^k)$ and $Var(\hat{\Omega}^k)$ ('plug-in' variance estimates) are biased upwards due to measurement error (Krueger & Summers (1988), Kline et al. (2020)). This will cause estimates of officer heterogeneity to be biased upwards. I bias-correct all (co)variance components using Kline et al. (2020)'s (KSS) heteroskedastic-consistent bias correction.[15]

Second, we are interested in the potential trade-offs between officer effects on one dimension relative to another, such as if officers who make more arrests are also those who commit costly misconduct or if the teams of officers who are best at preventing property crime also prevent violent crime. This is captured by the cross-outcome covariance, $Cov(\theta^k, \theta^{k'})$ and $Cov(\Omega^c, \Omega^{c'})$. Due to measurement error, 'plug-in' cross-outcome covariances are biased

[14] Bonhomme (2021) provides a pruning algorithm to iteratively remove indicators (e.g., officers) for whom effects cannot be identified, then removing any team in which they are on, until convergence.

[15] The leverages for the KSS correction for the officer-level analyses and alternative specifications use match-level (data aggregated to one row per officer-assignment after partialling out time-varying controls with proper weights) to avoid concerns over serial correlation in longer panels; non-match-level results are included in the robustness figures. While Bonhomme (2021) proposes a homoskedasticity robust correction and utilizes team-specific scaling factors in the team production model, in this environment the KSS assumption of the number of officers and number of teams tending to infinity is reasonable.

(downward), so I apply the Lachowska et al. (2023) extension of KSS to construct bias-corrected estimates of cross-outcome covariances.

Third, we are interested in the relationship between officer-activity effects and their influence on crime — e.g., whether officers who make more stops are best at reducing crime. However, because of the different levels of data (i.e., different design matrices), we cannot recover unbiased estimates for $Cov(\theta^k, \omega^c)$. One approach is to project officer-activity effects (shrunken to reduce measurement error),[16] denoted by $\tilde{\theta}_i^k$ to distinguish from unshrunken ('raw') estimates, where K is the set of officer-level activities, onto estimated officer-crime effects:

$$\hat{\theta}_i^c = \alpha_0 + \sum_{k \in K} \alpha_k \tilde{\theta}_i^k + \mu_i \tag{5}$$

While this will produce unbiased projection coefficients ($\hat{\alpha}_k$), if the degree of measurement error in $\hat{\theta}_i^c$ is large, then KSS-bias corrected heteroskedasticity robust standard errors will result in imprecise estimates. For increased precision, I directly estimate the causal effect of officer-activity effects on crime by regressing team-average (shrunken) officer-activity effects ($\overline{\theta}^k_{\tau(d,h)} = \frac{1}{N_{\tau(d,h)}} \sum_{i \in \tau(d,h)}^{N_{\tau(d,h)}} \tilde{\theta}_i^k$) on crime directly ($y^c_{d,h}$), effectively replacing high-dimensional exact team composition ($\Omega_{\tau d,h}$) in equation (3) with the low-dimensional team average across their activity effects:

$$y^c_{d,h} = \sum_{k \in K} \alpha_k \overline{\theta}^k_{\tau(d,h)} + \beta^c X_{d,h} + \gamma^c_{w(d,h)} + \epsilon^c_{d,h} \tag{6}$$

In this case, I use the ideal assignment-time controls ($\gamma^c_{w(d,h)}$, interacting unit, year, month, day of week, and hour fixed effects) because they can be removed through de-meaning and do not need to be estimated. Then $\hat{\alpha}_k$ tells us the strength of the relationship between officer-activity effect k and their effect on crime.

Fourth, we want to know the degree to which activity k directly influences crime. While α_k dictates the relationship between officers' effects on activity k and their effect on crime, it

[16] Effects are shrunken toward the mean as random intercepts to reduce measurement error using the mixed models method, similar to Rozema & Schanzenbach (2019).

does not speak to the direct effect. To estimate the direct effect of activity k on crime c (η_k^c), I estimate the two-stage least-squares regressions:

$$\begin{aligned} y_{d,h}^c &= \eta_k^c y_{d,h}^k + \beta^c X_{d,h} + \gamma_{w(d,h)}^c + \epsilon_{d,h}^c \\ y_{d,h}^k &= \rho_k^k \overline{\theta}_{\tau(d,h)}^k + \beta^k X_{d,h} + \gamma_{w(d,h)}^k + \epsilon_{d,h}^k \end{aligned} \tag{7}$$

where $y_{d,h}^k$ is the number of activity k (e.g., the number of stops made), and $y_{d,h}^c$ is the number of crimes. Because crime and policing are co-determined, I instrument for the endogenous variable $y_{d,h}^k$ using the team's average propensity for that activity $\overline{\theta}_{\tau(d,h)}^k$. Estimates of η_k^c can be used quantify elasticity of crime type c with respect to activity k, and, by applying costs to crimes and policing activities ($w^c \times y^c$ and $w^k \times y^k$), re-weighted estimates of η_k^c can determine if such activities produce net costs or net benefits (e.g., is the one additional street stop worth it?).

Fifth, and finally, we wish to know the change in total social costs from some change in officer behavior, such as replacing an officer with one who commits less misconduct. This involves applying social costs to crime (w^c) and officer activities (w^k) and using cross-activity covariances to determine the influence of such a change:

$$\frac{d \text{ Social Cost}}{d\theta^k} = \sum_{l \in K} [w^l + \sum_c w^c \times \eta_l^c] \times Cov(\theta^k, \theta^l) \tag{8}$$

where the total change in social costs from a change in one behavioral dimension (k) is the sum of how costs from all activities ($l \in K$) change due to covariance between officer effects plus the change in each crime type, weighted by its social cost, due to the change in each behavioral dimension.

5 Effect of Officers on Activity Outcomes

5.1 Officer Heterogeneity

Table 3 summarizes the influence of officer heterogeneity on activities after recovering officer fixed effects from estimating equation (1). The table displays the average number of officer activities (Column (1)) and bias-corrected standard deviation of officer effects ($SD(\theta_i^k)$) (Column (2)) per 100 shifts, along with the total amount of variation explained by officer and assignment effects (Column (3)), and the share of each component of the total variation in y^k (with low-dimensional controls, $x_{i,t}$, partialled out) explained by officer effects, assignment effects, and the sorting between the two (Columns (4)-(6)).[17]

Focusing on Columns (1) and (2), differences across individual officers are economically significant. Over 100 shifts, 1 SD increase in officer effects for each respective outcome increases street stops by 29.7 (36.4 pre-2016 and 15 post-2016) or 82.9% of the mean, traffic stops by 15.4 (14.3 pre-2016 and 27.1 post-2016) or 208% of the mean, nonindex arrests by 9.93 (88.4% of the mean), index arrests by 1.18 (73.6%), uses of force by 0.4 (109.8%), and accusations of misconduct by 0.199 (86.4%). Without the bias-correcting, these effects would appear noticeably larger, as is evidenced by the relatively large amount of bias in variance estimates in Figure C.1 (generally $> 20\%$ of $Var(\hat{\theta}_i^k)$).

We can compare some of these estimates to prior work, which takes the variance of officer effects shrunken using Empirical Bayes: Adger et al. (2022) finds a 1 SD increase in use of force propensity equates to a 138% increase compared with the mean among field training officers using 911 call data from Dallas, and Rozema & Schanzenbach (2019) finds a 1 SD increase in officer effects for civilian allegations of misconduct is around 140% of the mean. Both are larger than the estimates in this study but qualitatively similar. In contrast, Weisburst (2022) finds a 31% increase in arrests in response to a 1 SD increase in officer

[17]
$$Var(y_{i,t}^k - x_{i,t}\beta^k) = Var(\theta_i^k) + Var(\psi_{J(i,t)}^k) + 2Cov(\theta_i^k, \psi_{J(i,t)}^k) + Var(\epsilon_i^k)$$

effects using 911 call data from Dallas.

Using Columns (4) and (5), we can compare the degree of officer heterogeneity with that of heterogeneity in assignment effects. Recall that assignment effects encapsulate a wide range of different assignments and roles officers can work in. Despite this, the magnitudes of officer heterogeneity ($Var(\theta^k)$) tend to be as large as assignment effect heterogeneity ($Var(\psi_J^k)$). For street stops and nonindex arrest variance in officer effects is around 1.15 and 1.25 times the size of variance of assignment effects, while for traffic stops officer heterogeneity is nearly twice as large as assignment heterogeneity (1.79). Consistent with their less discretionary nature, officer heterogeneity for index arrests smaller than assignment heterogeneity, though $Var(\theta)$ is still around 0.49 the size of $Var(\psi)$. Officer heterogeneity is significantly larger than assignment heterogeneity for use of force (2.98), but less so for complaints (1.35), suggesting that differences across officers are a more significant factor in these highly costly outcomes compared with differences in assignments. Nevertheless, officer and assignment effects explain small portions of the overall variation in the rarer outcomes $< 1\%$ and around 2% for index arrests, though they explain a significant portion of the more common, highly discretionary activities such as stops and nonindex arrests (13-21% of overall variation).

Finally, officers sorting into assignments explains very little of the overall variation. In fact, correlations between officer and assignment effects are often negative or near zero for all outcomes other than nonindex arrests (which has positive correlations around $0.19 - 0.25$) and street stops to a lesser extent. This indicates that to the extent that officers sort into assignments that match their behavioral propensities, it is most consistent with sorting on opportunities for low-level arrests rather than highly rare outcomes such as force or misconduct. Furthermore, it rejects any notion that 'bad apples' (high force or high misconduct officers) sort into assignments where they can abuse their power, producing more misconduct and use more force. One implication is that raw use of force and complaint rates among officers, unconditional on their assignments, has a strong relationship with their causal effect on complaints and force.

5.2 Cross-Outcome Correlations

Given the importance of officer heterogeneity for each activity, we can determine the degree to which there is a trade-off between more 'productive' officers and officers who engage in costly behaviors based on how officer effects across different outcomes covary. Table 4 displays the KSS bias-corrected cross-outcome correlations in officer effects for each pair of outcomes and across sample periods. Interestingly, I do not find any very large (> 0.7) correlations between officer effects for any pair of outcomes. The strongest relationship is between the two measures of highly costly officer activity — complaints and use of force — with correlations between 0.54 and 0.63. This is followed by the correlations between arrest type (index and nonindex), which range from 0.42 to 0.64.

Furthermore, street and traffic stops, which are only weakly correlated pre-2016 (< 0.12), are correlated at 0.48 post-2016, with similar rises in the correlations between traffic stops index and nonindex arrests from < 0.06 and < 0.16 pre-2016 to ≈ 0.3 post-2016, while the correlations between nonindex arrests and street stops remained relatively stable between 0.35 to 0.45. This shift does suggest that as traffic stops overtook street stops in use, the set of officers who engaged in them most also shifted. However, the modest magnitudes in these correlations suggest that officers tend to be more specialized in specific behaviors.

Officer effects for complaints are at most modestly correlated with any outcome other than use of force (all < 0.35). Prior to 2016, complaints are negatively or weakly correlated with street and traffic stops (< 0.1), but in 2016-2019 they are weakly and positively correlated with both (0.16 and 0.12). Complaint effects also have a persistent but small positive correlation with serious (index) arrests, ranging from only 0.06 to 0.12. Even with nonindex arrests, which are often viewed as unnecessary and socially costly, complaint effects are only correlated at most 0.32.

Use of force effects follow a similar pattern but are more strongly correlated with all other outcomes. While correlations between use of force and street and traffic stops are still small (less than 0.3 and 0.14, respectively), correlations with arrests are modest in magnitude:

ranging from 0.37 to 0.41 for nonindex arrests, and from 0.32 to 0.39 for index arrests.

These relatively weak correlations between officer activity metrics and complaints and, to a lesser extent, use of force are highly inconsistent with a model of misconduct and force being an integral component of 'good' policing. Rather, the results are more consistent with use of force and especially misconduct being a costly behavior not integrally related to productivity. In Appendix D.1, I find that misconduct and use of force also do not have a significant relationship with officers' arrest quality, beyond their relationship with arrest quantity.

5.3 Robustness

I perform a number of robustness checks using alternative samples and specifications, which are described below. To summarize, the results of the robustness tests are consistent with the main results. Table B.1 displays sample sizes across most robustness specifications, and Figure C.2 tests officer effects' stability across these different samples by taking the correlation between officer effects from the main sample and each alternative sample. Even though these uncorrected correlations are biased toward zero, the cross-sample correlations are very strong, generally between 0.75-1 for most outcomes, with the lowest being complaints and use of force, though these are correlated with the main sample around 0.5-0.9 excluding a few outliers which cover small samples compared with the main specification.

Figure C.3 shows that standard deviations of officer effects (scaled by the mean of the outcome) are consistent across specifications. Figures C.4 and C.5 show that cross-outcome correlations in officer effects are also similar across specifications, with differences being noticeable only in samples that deviate significantly from the main sample. Cross-outcome correlations are generally smaller than in the main specification, suggesting weaker relationships between officer effects for costly and 'productive' activities.

I use three related alternative samples to overcome concerns over violations of exogenous mobility: I filter the data to unit-watch spells that last at least twenty shifts ("Long Unit-

Watch"), I filter the data to assignment spells that last at least four shifts ("Long GAB"), and I remove the first and last shift on each spell (i.e., removing the points just before and after the switches in Figure 2), which posed a concern due to pre-trends ("Donut"). Additionally, I take the sample of observations in which the officer is predicted to be working based on the rotations calendar to ensure endogenous selection of working days does not drive the results ("Predicted On").

Because the outcomes are count data with many zeros, I ensure the zero-inflated nature of the outcomes does not drive the results by filtering to a subset of assignments in which activity outcomes are more common (averaging over 0.25 activities per shift and with at least 300 observations) ("High Outcome"). Additionally, I estimate an analogous fixed effect regression using a Poisson specification ("Poisson") and use the Dhaene & Jochmans (2015) split-sample correction to de-bias estimates of θ_i ("Poisson Jackknife"), though this does reduce the number of officer and assignment effects recovered. Additionally, unlike the linear models, the (co)variance components for Poisson are not bias-corrected, so cross-outcome correlations are biased downward and/or attenuated.

The data also contain some officers working with partners, which may contaminate their individual officer effects. I ensure this does not drive the results by dropping all observations in which an officer is recorded as working in a car with another officer ("No Partner").

To determine if the results are influenced by the selection of sample periods, I use rolling 2-year sample periods from 2010 to 2019 ("Rolling 2-Year"). To ensure the results are consistent using the same filters as the district-hour analysis, I use only patrol units and regular watches 1-4 ("Regular Units/Watches"). To ensure the results are robust to using alternative definitions of assignments, I also change assignment definitions to be more (unit-watch-exact beat) ("Assign = Unit-Watch-Beat") and less stringent (unit-watch) ("Assign = Unit-Watch"). (Appendix D.2 compares main results to results using a complete match effects model.)

Finally, I aggregate the data to the officer-week level and select the most common

assignment within that week of working as their assignment ("Week"). This provides checks on multiple dimensions: mobility depends on officers working in different modal assignments between weeks, not just day-to-day; serial correlation is reduced; it allows me to check if officer effects are influenced by their usage of overtime. (Separately, Appendix D.3 shows that the ability for officers to extend their shifts using overtime does not prove to be a significant factor.)

6 Effect of Officers on Crime Outcomes

6.1 Team Heterogeneity

Table 5 summarizes the effect of team heterogeneity (a function of officer heterogeneity) on the number of crimes reported using the first-difference estimator in equation (4). Variation in the composition of a team of officers explains a modest but economically meaningful amount of the variation in reported crimes. For index crimes, variation in team composition explains between 0.39-1.05% of the overall variation in property crime but only 0.16 - 0.2% of violent crime. For comparison, variation in team composition explains about 0.51-0.67% of variation in reported drug crimes. Given that drug crimes are entirely drug arrests and officer-level heterogeneity explains a significant portion of nonindex arrests as shown in Section 5.1, the explanatory power of team effects on property crime and violent crime is significant.

The magnitude of team heterogeneity in crime effects is also economically significant. For violent crime, a 1 SD increase in team effects (a team that is worse at preventing crime) results in about 0.02 more crime per hour, or around a 15% increase relative to the mean, with similar effects on both robbery and aggravated assaults. For property crime, a 1 SD increase in team effects results between 0.06 and 0.11 additional crimes per hour, or between a 13-20% increase, primarily driven by effects on theft.

To better quantify the magnitude of these effects, we can focus on how a 1 SD better (crime-reducing) officer changes crime outcomes compared with adding a single officer to a

team. As much of the crime literature studies the elasticity between the quantity of officers and crime, I use elasticity estimates from Mello (2019) and Chalfin & McCrary (2018) (CM) (which includes standard errors) to compute these quantity effects of adding a single *average* officer to the average team — going from 49 to 50 officers, or increasing officer quantity by 2%. I compare this to the effect of holding team size fixed but increasing a single officer by 1 SD in terms of crime-reduction quality. Figure 3 compares the influence of the quantity and the 'quality' of officers. A 1 SD increase in quality effects decreases violent crime by 0.3%, around half the size of CM (0.6%) and Mello (2.5%) quantity effects. For property crime, a 1 SD increase in quality reduces crimes by around 0.5%, the same as the CM estimates but smaller than those of Mello (1.6%). Overall, quality effects are comparable but smaller than implied quantity effects for violent crimes and similarly sized for property crimes.

6.2 Mechanisms: Deterrence, Incapacitation, Suppression

There are three primary mechanisms through which officers can influence reported crimes. First, officers may reduce crime through contemporaneous deterrence (Mello (2019), Chalfin et al. (2022)) through higher quality positioning, tactics, and visibility to increase the perceived probability of detection. Alternatively, officers may reduce crime by arresting and thus incapacitating offenders who would otherwise commit (additional) crimes (Mastrobuoni (2019)). Finally, officers may influence crime rates by differentially suppressing crime reports rather than influencing underlying victimization rates.

If officers are reducing crime through arrests and incapacitation, we would expect that teams that tend to make more arrests should also reduce unsolved crimes, implying a negative covariance between team effects for unsolved crimes and arrests for the same crime types (if criminals broadly specialize within crime types). Specifically, because Crime = Arrests+Uncleared Crimes, the variance of team effects on any crime type can be decomposed into: $Var(\Omega^{Crime}) = Var(\Omega^{A}) + Var(\Omega^{U}) + 2 \times Cov(\Omega^{A}, \Omega^{U})$.

Figure C.6 displays these components normalized by the overall variance of the crime

effects. The primary takeaway is that team effects for uncleared crime and arrests are generally positively correlated, though the covariance between the two is relatively negligible compared with the variance in arrests for drug crimes and the variance in uncleared offenses for property crime. This suggests that officers influence property crime by influencing uncleared offenses rather than arrests. Additionally, officers who make more arrests of a given type do not also reduce uncleared crimes, which is inconsistent with incapacitation and more consistent with officers being differentially effective at deterring crime.

To determine if officers are simply influencing the reporting of crime rather than crime itself, I use 911 calls as a proxy for underlying crime that officers cannot easily manipulate. Figure C.7 displays SDs of team effects scaled by the mean on 911 calls of different types and comparable crime types. The influence of teams on 911 calls is similar to that of teams on related crime types (e.g., calls about theft and reported thefts, assaults, robberies, etc.). And, as discussed in Appendix D.5, team effects on 911 calls are correlated with team effects on crime reports of related types. This is consistent with officers having a differential ability to prevent crime rather than manipulating crime reporting through suppression.

6.3 Robustness

I employ two main robustness tests for this analysis. Table C.3 displays the results, and Appendix D.4 discusses alternative team constructions. First, I directly exploit the calendar rotation by using the predictions of which officers are working in which units on which days and predict their start/end times. With these officer-level predictions of who is working in a given district-hour, I construct a predicted team and use that as the team composition ($\Omega_{\tau(d,h)}$) in the main first-differences ('FD') specification ('PFD'). This recovers the effects of ω_i based on variation in the predicted starting/stopping work in a given unit and hour, analogous to the reduced form regression of instrumenting for potentially endogenous changes in team composition using exogenous predicted changes in team composition. The estimates are generally consistent with the main specification and calendar instrument's strong predictive

power. Notably, the PFD variance estimates on drug crimes are around twice the size of the FD ones, consistent with officers impacting discretionary arrests more strongly than property and violent crimes, though those influences are still economically significant.

Second, I take the fixed effects ('FE') specification from equation (3) and use feasible assignment-time fixed effects ($\gamma_{w(d,h)}$) by interacting year, month, district, day of week, and 8 hour period (as opposed to exact hour as in the ideal specification). The shares of variance explained are much larger in FE than in FD, consistent with the feasible assignment-time effects being insufficiently granular.[18] Nevertheless, the relative magnitudes of officer influence on crimes are similar: officer effects explain much more variation in drug and property crime than in violent crime, though, as with PFD, variation explained for drug crime is around twice that of property crime.

7 Connecting Policing and Crime

7.1 Effect of Officer-Activity Effects on Crime

To determine which officers are better at reducing crime, I connect officers' effects on activities (stops, arrests, misconduct, etc.) to their influence on crime. Table 6 displays the results of projecting officer-crime effects onto shrunken officer-activity effects (equation (5)), and Table 7 displays the results of regressing crimes at the district-hour level onto average shrunken officer-activity effects of officers working on that team (district-hour) (equation (6)). In both cases, the coefficients correspond to a 1 SD increase in the independent variable divided by the mean of the initial outcome (e.g., property crimes per district-hour) for comparability across specifications and outcomes.

The crime regression results are much more precise than the projection coefficients, consistent with the sizable amount of measurement error in the officer-crime effect estimates.

[18] Additionally because the number of parameters to bias correct goes from around 10,000 in first differences to nearly 30,000 in fixed effects (due to including $\gamma_{w(d,h)}$'s in addition to all officer effects), some KSS variance components are negative.

Even with a few statistically significant coefficients for the projection results, there are no directionally consistent effects across sample periods, and complaints and use of force effects have no negative and statistically significant effects on violent or property crime. With the regression approach, I find that teams with higher propensities to make drug arrests have large, positive, and statistically significant effects across all sample periods, as expected. For violent and property crimes, however, I find few statistically significant or economically significant effects. In the pooled samples, use of force and complaint activity effects have very small and near zero effects estimated with relative precision, ruling out effects larger than +2% to -2% for violent crime and larger than -0.4% to 1.2% for property crime per 1 standard deviation in either activity effect. Among stops and arrests activity effects in the pooled sample, only non-drug nonindex effects have a statistically significant effect, with a 1 SD increase in team average non-drug nonindex arrest effects decreasing property crime by 1% of the mean ($p < 0.05$).

To probe the robustness of these results, I avoid any concerns relating to officers endogenously selecting to work on a given day in the regressions of crime on team activity propensities by instrumenting for the actual team propensities using predicted team activity propensity. I use the calendar rotation to predict who is working in a given district-hour, denoting this as 'IV' compared with the main specification 'OLS'. I also use unshrunken rather than shrunken officer-activity effects. Figure C.8 displays the comparison, and the interpretation is largely unchanged, though instrumenting for team composition tends to produce larger and more noisy estimates.

7.2 Effect of Policing Activities on Crime

Given this weak relationship between officer-activity effects and their effect on crime, I turn to the fourth set of parameters of interest: the effect of policing activity on crime. For this, I regress crime on a specific policing activity (e.g., total number of stops in a district-hour) and use team average officer-activity effects as the instrument for total activities (equation

(7)). Panel B of Table C.4 shows the first stage F-statistics for the relationship between total activities and team average propensities for those activities, which are strong for the activities of interest such as street and traffic stops and nonindex arrests.

I convert the estimates from estimating equation (7) from the effect of 1 additional activity of type k on crimes of type c per district-hour to the elasticity of crime c with respect to activity k by scaling coefficients by the respective means. Table 8 displays the results for property and violent crime in the pooled sample for officer-activities street and traffic stops, index arrests, nonindex arrests, and the most discretionary types of arrests for the lowest levels of crime, 'quality-of-life' (QOL) (Cho et al. (2023)).

For violent crime, I find no statistically or economically significant elasticity across all arrest types (index, nonindex, QOL) and stop types (street and traffic). Though not statistically significant, the effects are precisely estimated enough that I can reject elasticities of violent crime with respect to arrests larger than -0.16, with point estimates smaller than -0.07. For stops, point estimates are even smaller with the elasticities with respect to street and traffic stops being -0.019 and -0.013, respectively, and 95% confidence intervals rejecting elasticities larger than -0.084.

For property crime, I find small but statistically significant elasticities with respect to nonindex arrests, QOL arrests, and street stops (all $p < 0.01$), but near zero effects for traffic stops. While the elasticity of property crime with respect to index arrests is -0.02, with respect to nonindex arrests the elasticity is -0.1. Interestingly, the elasticity of property crime with respect to arrests is not driven by arrests for QOL offenses, as the elasticity is much smaller than that of general nonindex arrests, at -0.062. Similarly, street stops have a small but precisely estimated elasticity of -0.053. Nevertheless, these are small effects, with 95% confidence intervals rejecting elasticities larger than -0.16 for index and nonindex arrests, larger than -0.1 for QOL arrests, and larger than -0.1 for street stops.

We can compare these lower bounds to the smallest across recent estimates for the elasticity of crime with respect to police quantity of -0.34 for violent and -0.174 for property

crime from Chalfin & McCrary (2018), which are generally larger in magnitude than even the lower bounds of the confidence intervals around the elasticities of crime with respect to policing activities. Notably, the Chalfin & McCrary (2018) elasticities are, themselves, significantly smaller than other relevant estimates from Mello (2019), Weisburst (2019), Evans & Owens (2007), and Lin (2009). Furthermore, Cho et al. (2023) calculates implied elasticities of violent and property crimes with respect to arrests at 0.42 and -0.15, but with confidence intervals that cannot reject elasticities between -0.4 and 0.3 for either. However, I can also reject even small positive elasticities (larger than +0.1) between crime and any of these officer activities. Overall, I find relatively small and precise elasticities of crime with respect to policing activities.

Table 8 also displays three robustness tests alongside the main results. First, I use unshrunken officer effects as the instrument with similar results, though the elasticity of violent crimes with respect to QOL arrests is twice as large (-0.076, $p < 0.05$). Second, I use the predicted team composition as the instrument rather than actual team composition, which results in weaker first stages and larger elasticities for violent crimes (though all $p > 0.05$) and small and not statistically significant elasticities for property crime. Finally, there is potential for exclusion restriction violations because, for example, officers' propensities to make stops may also increase use of force, which then influences crime. So, I borrow from the judge-IV literature and, in an additional specification, I add the predicted team average officer effects for the other officer activities $\overline{\hat{\theta}^{\setminus k}}^{p}_{d,h}$ as controls (Mueller-Smith (2015)) — for example, I use propensities for street stops as the instrument for total stops and control for arrest, use of force, and other activity propensities. The first stages remain strong, and the point estimates are not significantly different from the main results. Overall, the results are consistent, though tests lean toward even smaller elasticities for property crime when using predicted teams.

8 Quantifying the Trade-Off

Finally, I combine estimates from the prior sections with social cost estimates for crime and policing activities to quantify the extent of a trade-off between aggressive policing and crime reduction following the logic of equation (8). I calculate how aggregate activity, crime outcomes, and total costs would change if we replaced a 'bad' cop with less aggressive officer officer, specifically an officer with 1 SD lower complaint or use of force. As shown in Panel A of Table 9, over 100 shifts, an officer with a 1 SD lower officer effect for complaints (use of force) will cause between 0.06 and 0.4 (0.34 and 0.44) fewer complaints (uses of force). However, other policing activities also change due to the cross-outcome covariances in officer effects, based on the cross-outcome correlations in officer effects in Table 4.

I apply social costs to each of these activities using conservative cost estimates. I value a use of force by an officer at $23,864, equivalent to the cost of a criminal assault excluding criminal-related costs (Miller et al. (2021)). I value a single complaint at $122,390, which is 40% of the lower bound of the 95% confidence interval for the relationship between officer complaints and lawsuit settlement costs (see Appendix D.6). I value a nonindex arrest at $13,785, which is 25% of the lower-bound estimate from Dobbie et al. (2018) for the net cost of pretrial detention relative to release. I value index arrests at $0, conservatively assuming that they prevent enough crime through incapacitation and general deterrence to be net-zero. Similarly, I value traffic and street stops at a lower bound of $0.[19] Despite the conservative cost estimates for these activities, I find significant cost savings from these 1 SD shifts: 1 SD decline in officer complaint (use of force) effects saves between $75,950 ($81,740) and $35,640 ($53,780) over 100 shifts. In terms of composition, the effects are driven in similar parts by

[19]Existing work on pretrial detention, prosecution of low-level offenses, and incarceration find significant net costs, lower earnings, higher future convictions, and relatively small incapacitation or general deterrent effects. See Dobbie et al. (2018), Mueller-Smith (2015), Agan et al. (2023), Humphries et al. (2023), Garin et al. (2023), and Kamat et al. (2023). Norris et al. (2021) finds crime-reducing effects on children from parental incarceration, though they are unable to reject net social costs. These estimates are conservative for various reasons. For example, use of force costs ignores spillovers on education (Ang et al. (2021)) and civic participation (Ang & Tebes (2023)); stops incur costs in the forms of individuals' time, stress, and potentially result in ticketing or fines (Mello (2021)).

the higher costs but smaller changes in complaints and use of force and the lower costs but higher frequency of nonindex arrests.

While gains from reducing activities are large, the primary cost associated with reducing officer complaint (use of force) effects is not due to their direct effect on crime. Rather, it is due to the correlation between complaint (use of force) effects and arrest and stop effects, which have larger effects on crime. To quantify the trade-off and following equation (8), I apply the declines in arrests and stops from 1 SD decline in an officer's complaint (use of force) effects over 100 shifts to the coefficients estimated from equation (7). I apply social costs to crimes based on Miller et al. (2021), with the values for each crime type displayed in Table C.5. Because murder is both rare and extremely costly — around $8 million (similar to 200 robberies) — I also provide a 'low murder' (LM) specification, valuing murder at $400,000 to avoid skewing results with rare murders, similar to the procedure in Heckman et al. (2010). Table C.6 displays the effects of each policing activity in Table 8 on crime costs.

Despite the conservative estimates used for activity costs, the results suggest that savings from reducing complaint and use of force effects well exceed costs from increased crime. As shown in Panel B of Table 9, over 100 9-hour shifts, the resulting decline in policing activities has highly imprecise effects using full murder costs. Comparison with the LM cost results suggests that this imprecision is due to rare but extremely costly murder effects. Using LM cost, estimates become significantly smaller and more precise, though most are still not statistically significant. Summing over the crime cost estimates for each dimension of officer activity (Panel C), I find that a 1 SD decline in complaint (use of force) effects over 100 shifts increases crime costs by between $55,670 ($232,420) and $16,870 ($29,240) using the imprecise full murder costs and between $25,230 ($40,080) and -$12,150 (-$25,210) using LM costs. Except for a single full murder cost estimate, the savings from activity costs exceed the costs from increased crime by at least $9,000.

For an additional set of estimates, I also use the coefficients from regressing crime costs on average team effects (equation (6)) and apply the single officer 1 SD change (changing

average team effects for one officer on an average team) and sum over the effects. As shown in Panel D of Table 9, while three full murder cost estimates exceed the activity cost savings, this is entirely driven by outlier murder costs as all of the LM cost estimates are significantly smaller than the activity cost savings.

9 Conclusion

This paper studies the trade-off between officers' propensity to engage in costly behaviors, such as using force and misconduct, enforcing the law, and reducing crime. While officer heterogeneity is significant across outcomes, I find that officers who engage in the most costly activities do not generate a large enough reduction in crime to outweigh the social costs of behaviors. In short, 'bad cops' are not better police for two reasons. First, officers' causal effects on use of force and misconduct are strongly correlated with each other but modestly or weakly correlated with enforcement activities (stop and arrests). These findings directly contradict a model of policing in which the most productive officers are also those who engage in overly aggressive policing. It suggests that such behaviors are not signals of effort, and curtailing through training or oversight is not likely to pose a direct threat to police effectiveness. Second, the direct effects of policing activities on crime are relatively small. As a result, reducing misconduct and use of force produces modest second-order effects on enforcement activities, which themselves have only modest effects on crime.

While these results suggest that there is little trade-off in training or recruiting less costly police officers in terms of increased crime or fewer important arrests, they also suggest that the features that make an officer good at preventing crime are not clearly linked to traditional metrics of officer activity. Determining which features predict effective crime fighting in officers and which training and recruitment strategies can produce good cops and better policing is ground for future work.

10 References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. *Longitudinal Employer-Household Dynamics Technical Papers*.

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333.

ACLU. (2015). *Stop and frisk in chicago*. ACLU of Illinois.

Adger, C., Ross, M., & Sloan, C. (2022). *The effect of field training officers on police use of force* (p. 57) [Working Paper].

Agan, A., Doleac, J. L., & Harvey, A. (2023). Misdemeanor prosecution. *The Quarterly Journal of Economics*.

Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *171*(3), 673–697.

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

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2021). Police violence reduces civilian cooperation and engagement with law enforcement. *SSRN Electronic Journal*.

Ang, D., & Tebes, J. (2023). Civic responses to police violence. *American Political Science Review*, 1–16.

Ba, B., Bayer, P., Rim, N., Rivera, R., & Sidibé, M. (2021). *Police officer assignment and neighborhood crime* (Working Paper 29243). National Bureau of Economic Research.

Ba, B., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in chicago. *Science*, *371*(6530), 696–702.

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

Best, M. C., Hjort, J., & Szakonyi, D. (2023). Individuals and organizations as sources of

state effectiveness. *American Economic Review*, *113*(8), 2121–2167.

Bonhomme, S. (2021). *Teams: Heterogeneity, sorting, and complementarity* (arXiv:2102.01802). arXiv. http://arxiv.org/abs/2102.01802

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., & Setzler, B. (2022). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*.

Campbell, R. A. (2024). *What does federal oversight do to policing and public safety? Evidence from seattle* [Working Paper].

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015.

Chalfin, A., Hansen, B., Weisburst, E. K., & Williams Jr., M. C. (2022). Police force size and civilian race. *American Economic Review: Insights*, *4*(2), 139–158.

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

Chalfin, A., & McCrary, J. (2018). Are u.s. Cities underpoliced? Theory and evidence. *The Review of Economics and Statistics*, *100*(1), 167–186.

Chan, D. C. (2021). Influence and information in team decisions: Evidence from medical residency. *American Economic Journal: Economic Policy*, *13*(1), 106–137.

Chen, Y. (2021). Team-specific human capital and team performance: Evidence from doctors. *American Economic Review*, *111*(12), 3923–3962.

Cheng, C., & Long, W. (2018). Improving police services: Evidence from the french quarter task force. *Journal of Public Economics*, *164*, 1–18.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Cho, S., Gonçalves, F., & Weisburst, E. (2023). *The impact of fear on police behavior and public safety* (31392). National Bureau of Economic Research. https://www.nber.org/p

apers/w31392

Corman, H., & Mocan, N. (2005). Carrots, sticks, and broken windows. *The Journal of Law and Economics*, *48*(1), 235–266.

Cox, R., & Cunningham, J. P. (2021). Financing the war on drugs: The impact of law enforcement grants on racial disparities in drug arrests. *Journal of Policy Analysis and Management*, *40*(1), 191–224.

Denman, D. S. (2023). The cinematic universe of copaganda: World-building and the enchantments of policing. *Culture, Theory and Critique*, *0*(0), 1–16.

Devi, T., & Fryer, J., Roland G. (2020). *Policing the police: The impact of "pattern-or-practice" investigations on crime* (27324). National Bureau of Economic Research. https://www.nber.org/papers/w27324

Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, *82*(3), 991–1030.

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

Dobbie, W., Goldin, J., & Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, *108*(2), 201–240.

Donohue, J. J., Ho, D. E., & Leahy, P. (2013). Do police reduce crime?: A reexamination of a natural experiment. In *Empirical legal analysis*. Routledge.

Draca, M., Machin, S., & Witt, R. (2011). Panic on the streets of london: Police, crime, and the july 2005 terror attacks. *American Economic Review*, *101*(5), 2157–2181.

Dube, O., MacArthur, S. J., & Shah, A. K. (2023). *A cognitive view of policing* (31651). National Bureau of Economic Research. https://www.nber.org/papers/w31651

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

Feigenberg, B., & Miller, C. (2022). Would eliminating racial disparities in motor vehicle searches have efficiency costs? *The Quarterly Journal of Economics*, *137*(1), 49–113.

Frederiksen, A., Kahn, L. B., & Lange, F. (2020). Supervisors and performance management systems. *Journal of Political Economy*, *128*(6), 2123–2187.

Garin, A., Koustas, D., McPherson, C., Norris, S., Pecenco, M., Rose, E. K., Shem-Tov, Y., & Weaver, J. (2023). *The impact of incarceration on employment, earnings, and tax filing* [Working Paper]. Working Paper.

Goncalves, F., & Mello, S. (2021). A few bad apples? Racial bias in policing. *American Economic Review*, *111*(5), 1406–1441.

Gudgeon, M., Jordan, A., & Kim, T. (2023). *Do teams perform differently under black and hispanic leaders? Evidence from the chicago police department* [Working Paper]. Working Paper.

Hausman, D., & Kronick, D. (2023). The illusory end of stop and frisk in chicago? *Science Advances*, *9*(39).

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope perry preschool program. *Journal of Public Economics*, *94*(1), 114–128.

Hoekstra, M., & Sloan, C. (2022). Does race matter for police use of force? Evidence from 911 calls. *American Economic Review*, *112*(3), 827–860.

Holz, J. E., Rivera, R., & Ba, B. A. (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy*, *15*(2), 256–291.

Humphries, J. E., Ouss, A., Stavreva, K., Stevenson, M. T., & Dijk, W. van. (2023). *Conviction, incarceration, and recidivism: Understanding the revolving door.*

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4), 1034–1073.

JG0304. (2009, November 18). *New CPD schedule! Officer.com.*

Jochmans, K., & Weidner, M. (2019). Fixed-effect regressions on network data. *Econometrica*, *87*(5), 1543–1560.

Kamat, V., Norris, S., & Pecenco, M. (2023). *Conviction, incarceration, and policy effects in the criminal justice system* [Working Paper].

Kelling, G. L., & Wilson, J. Q. (1982). Broken windows. *The Atlantic*.

Kim, T. (2024). *Promotion incentives, career decisions, and police performance* [Working Paper]. Working Paper.

Klick, J., & Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. *The Journal of Law and Economics*, *48*(1), 267–279.

Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, *88*(5), 1859–1898.

Krueger, A. B., & Summers, L. H. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica*, *56*(2), 259.

Lachowska, M., Mas, A., Saggio, R., & Woodbury, S. A. (2023). *Work hours mismatch* (31205). National Bureau of Economic Research. https://www.nber.org/papers/w31205

Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: reply. *American Economic Review*, *92*(4), 1244–1250.

Lin, M.-J. (2009). More police, less crime: Evidence from US state data. *International Review of Law and Economics*, *29*(2), 73–80.

Linnemann, T. (2022). *The horror of police*. University of Minnesota Press.

Mac Donald, H. (2016). *The war on cops : How the new attack on law and order makes everyone less safe* (First American Edition.). New York: Encounter Books.

Markovic, J. (2009). Why neighborhoods matter: The importance of geographic composition. *Geography & Public Safety*, *2*.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics*, *121*(3), 783–821.

Mastrobuoni, G. (2019). Police disruption and performance: Evidence from recurrent

redeployments within a city. *Journal of Public Economics*, *176*, 18–31.

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

Mello, S. (2021). *Fines and financial wellbeing* (p. 99) [Working Paper].

Mikdash, M., & Zaiour, R. (2023). *The impact of police shootings on gun violence and civilian cooperation* [Preprint]. SSRN.

Miller, T. R., Cohen, M. A., Swedler, D. I., Ali, B., & Hendrie, D. V. (2021). Incidence and costs of personal and property crimes in the USA, 2017. *Journal of Benefit-Cost Analysis*, *12*(1), 24–54.

Moser, W. (2012). Garry McCarthy's new chicago crime strategy: Social networks, 'hot people'. *Chicago Magazine*.

Moskos, P. (2008). Under arrest: Discretion in the ghetto. In *Cop in the hood* (STU - Student edition, pp. 111–157). Princeton University Press.

Mueller-Smith, M. (2015). *The criminal and labor market impacts of incarceration* (p. 59) [Working Paper].

Norris, S., Pecenco, M., & Weaver, J. (2021). The effects of parental and sibling incarceration: Evidence from ohio. *American Economic Review*, *111*(9), 2926–2963.

Ouss, A., & Rappaport, J. (2020). Is police behavior getting worse? Data selection and the measurement of policing harms. *The Journal of Legal Studies*, *49*(1), 153–198.

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2018). Can you build a better cop? *Criminology & Public Policy*, *17*(1), 41–87.

Parker, S. T., Ross, M. B., & Ross, S. (2024). *Driving change: Evaluating connecticut's collaborative approach to reducing racial disparities in policing* (32692). National Bureau of Economic Research. https://www.nber.org/papers/w32692

Premkumar, D. (2022). *Public scrutiny, police behavior, and crime consequences: Evidence from high-profile police killings* (p. 132) [Working Paper].

Prendergast, C. (2001). *Selection and oversight in the public sector, with the los angeles police department as an example* (w8664). National Bureau of Economic Research.

Prendergast, C. (2007). The motivation and bias of bureaucrats. *The American Economic Review; Nashville*, *97*(1), 180–196.

Rivera, R. (2022). *The effect of minority peers on future arrest quantity and quality* (SSRN Scholarly Paper 4067011). Social Science Research Network.

Rivera, R., & Ba, B. A. (2023). The effect of police oversight on crime and misconduct allegations: Evidence from chicago. *The Review of Economics and Statistics*, 1–45.

Rose, E. K., Schellenberg, J. T., & Shem-Tov, Y. (2022). *The effects of teacher quality on adult criminal justice contact* (30274). National Bureau of Economic Research. https://www.nber.org/papers/w30274

Rozema, K., & Schanzenbach, M. (2019). Good cop, bad cop: Using civilian allegations to predict police misconduct. *American Economic Journal: Economic Policy*, *11*(2), 225–268.

Shi, L. (2009). The limit of oversight in policing: Evidence from the 2001 cincinnati riot. *Journal of Public Economics*, *93*(1), 99–113.

Skogan, W. G. (2022). Prospects for reform? The collapse of community policing in chicago. *The University of Chicago Law Review*, *89*(2), 383–404.

Song, J., Price, D. J., Guvenen, F., Bloom, N., & Wachter, T. von. (2019). Firming up inequality. *The Quarterly Journal of Economics*, *134*(1), 1–50.

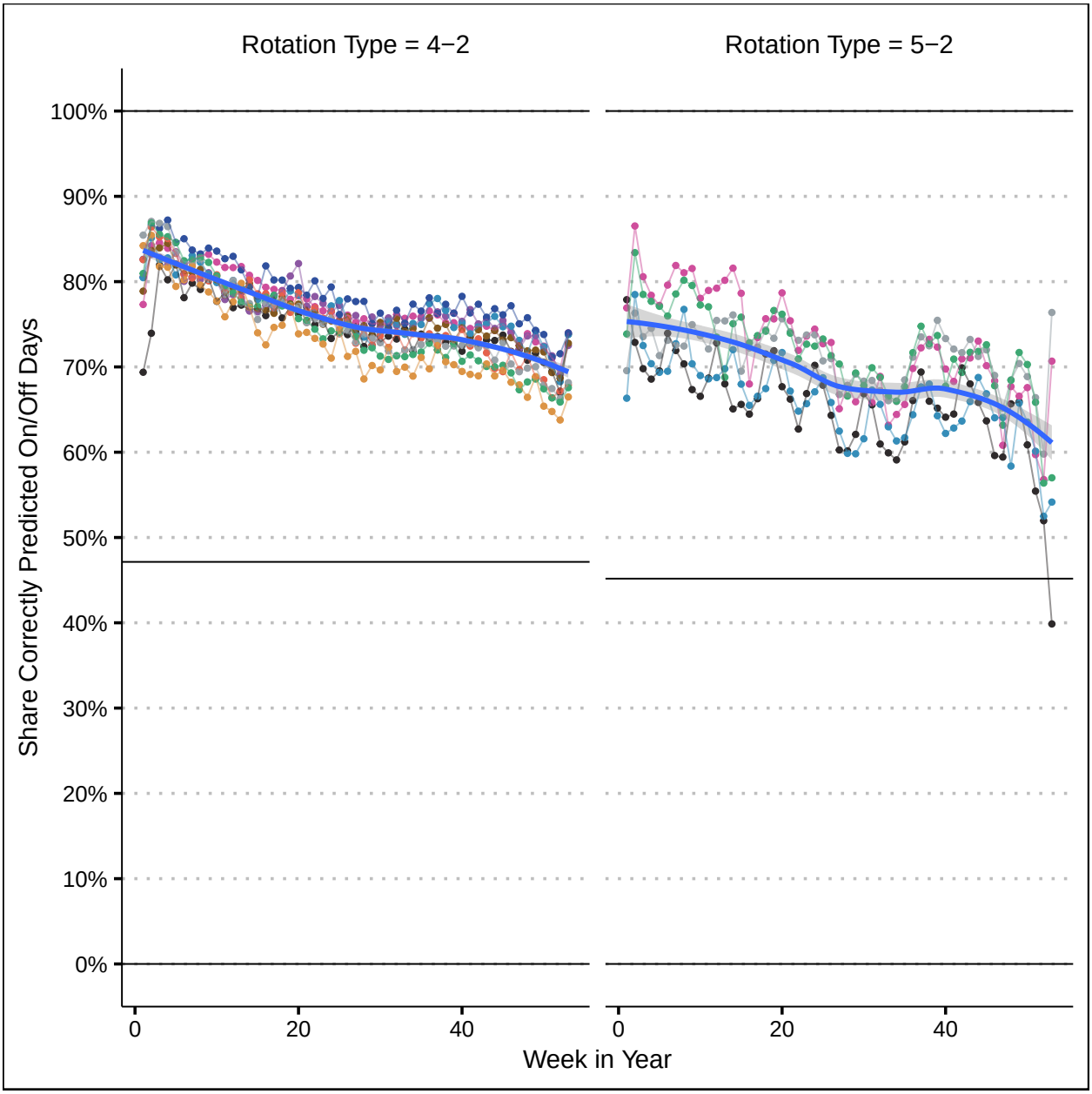
Tebes, J., & Fagan, J. (2025). *Do pedestrian stops deter crime? Evidence from reforming "stop and frisk"* [Working Paper]. Working Paper.

Walters, C. R. (2024). *Empirical bayes methods in labor economics* (33091). National Bureau of Economic Research. https://www.nber.org/papers/w33091

Weisburst, E. K. (2019). Safety in police numbers: Evidence of police effectiveness from federal COPS grant applications. *American Law and Economics Review*, *21*(1), 81–109.

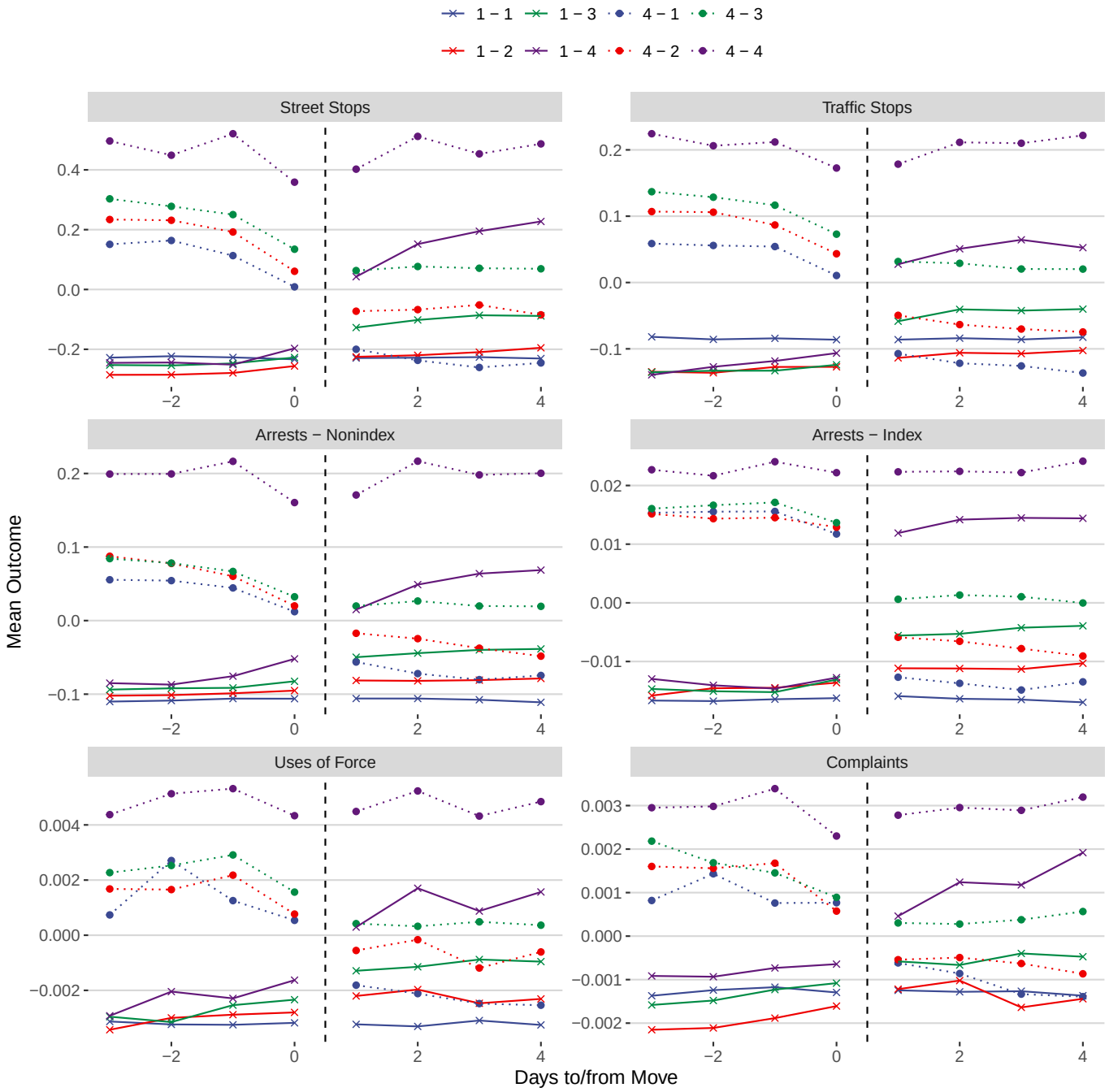
Weisburst, E. K. (2022). "Whose help is on the way?": The importance of individual police officers in law enforcement outcomes. *Journal of Human Resources*.

Figure 1: Accuracy of Predicted Days On and Off



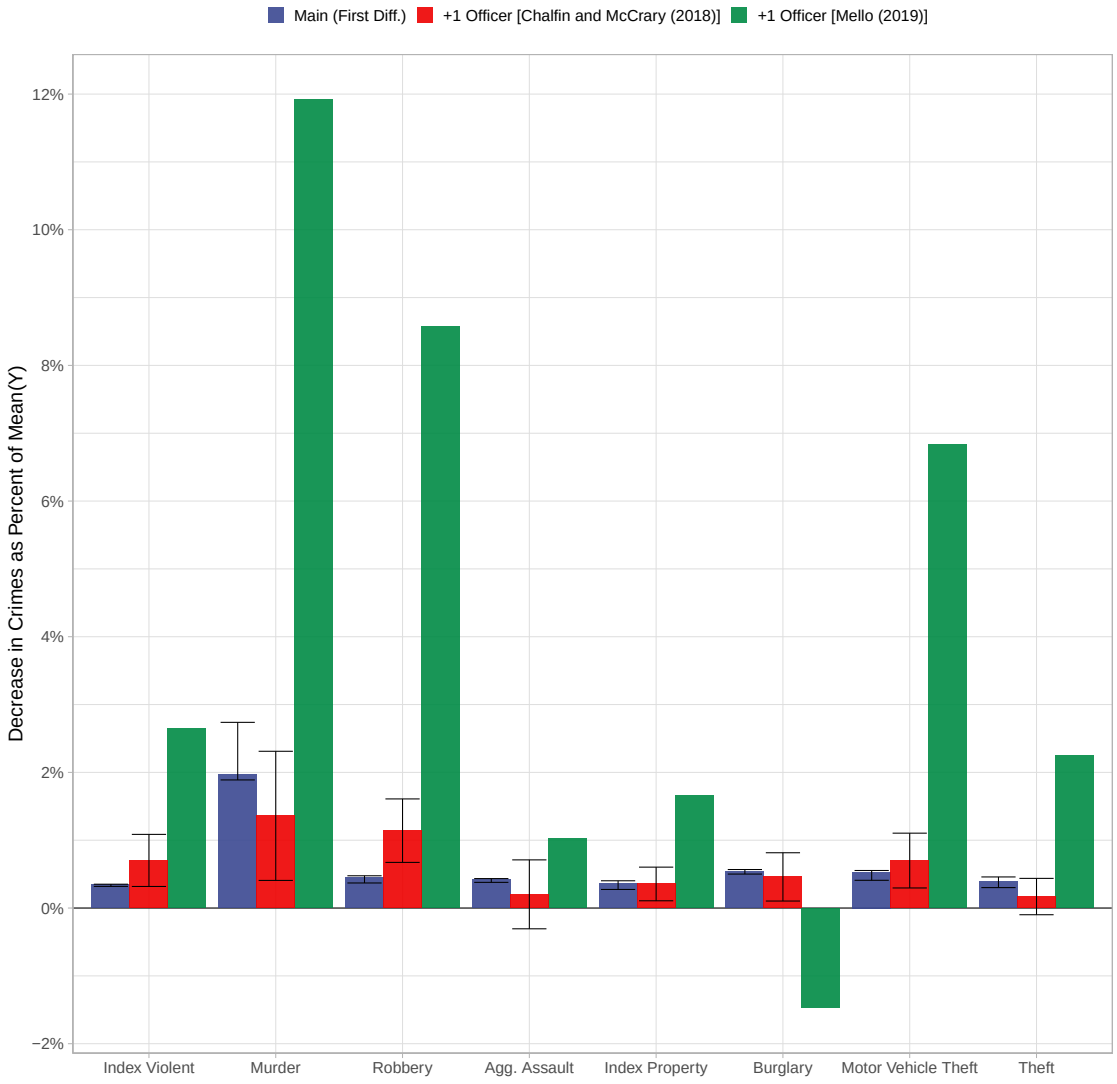
Note: Figure displays the share of correctly predicted days-on and days-off using based on projecting backward and forward the first 4 days-on and 2 days-off (4-2) and 5-2 spells, with average correct predictions aggregated at the year-week level and each color denoting a different year of data. The horizontal line corresponds to the accuracy of a random set of days given the spell duration.

Figure 2: Mean Outcomes of Assignment Switchers by Quantile of Mean Outcome in Assignment



Note: Figure displays the mean outcome for officers moving between assignments (unit-watch-beat type), with outcomes demeaned by year, grouped by the quantile of the assignment they are transitioning to and from, in the seven shifts leading up to and after the switch (x-axis), with the shift right before the switch being 0. The quantile of an officer's assignment is computed at the outcome-year-level and leaving out the officer themselves, and quantiles of the leave-out values are weighted by observations within each year. Figure only displays transitions for officers in quantile 1 (lowest), with star shapes and solid lines, and quantile 4 (highest), with circle shapes and dotted lines, outcome assignments prior to their switch, while colors correspond to the quantile of the assignment the officer is switching to. Not all officer-assignment spells are six shifts or longer, so the set of officer-assignment spells contributing to the mean outcomes for x=1 is larger than the set of officer-assignment spells contributing to the mean outcomes for x=2, for example.

Figure 3: Effect of Officer Quantity versus Quality



Note: Figure displays comparison between the effect of 1 additional officer added to a team using elasticities of crime with respect to number of officers from estimates from Mello (2019) and Chalfin & McCrary (2018) across index crime types ('quantity' effect) compared with the effect of reducing a single average officer's crime effects by 1 SD as recovered in this paper ('quality' effect). Elasticity estimates are converted to percents of the mean by multiplying the elasticity estimates by 1/49 (the percent increase in number of officers by adding 1 officer to the average team of 49 officers). Quality estimates are computed from the KSS bias corrected standard deviations of team effects (Ω), recovered from the main first-differences design, with each estimate being divided by 49 (the average team size) and divided by the mean of the outcome to provide an approximation of the effect of 1 SD decrease in a single officer's crime effect on an average team. Each bar for the non-quantity estimates represents the median of estimates recovered from each of the three multi-year sample periods. Note that if the covariance between officer effects on teams is relatively large in magnitude compared with the variance across officers, then this is a poor approximation. Confidence intervals on the Chalfin & McCrary estimates correspond to converted 95% CIs using standard errors reported in the paper, and confidence intervals on the quality effects denote maximum and minimum values across the samples/specifications.

Table 1: Summary Statistics of Officer-Level Data

	Pooled	2010 - 2012	2013 - 2015	2016 - 2019
	(1)	(2)	(3)	(4)
A: By Officer				
Avg. Total Assignments	14.1 (8.56)	7.6 (4.95)	7.89 (5.01)	8.86 (5.49)
Avg. Total Units	2.12 (1.3)	1.55 (0.86)	1.39 (0.692)	1.49 (0.776)
Avg. Share Most Common Assignment	0.477 (0.0975)	0.615 (0.147)	0.594 (0.148)	0.597 (0.145)
Avg. Share Most Common Unit	0.849 (0.216)	0.923 (0.186)	0.938 (0.174)	0.916 (0.193)
Avg. Obs. (Shifts)	1,373 (526)	522 (114)	517 (126)	612 (209)
Geographic Unit	0.79 (0.407)	0.793 (0.405)	0.794 (0.405)	0.785 (0.411)
Movers	0.997 (0.051)	0.989 (0.106)	0.979 (0.142)	0.983 (0.13)
B: Pooled				
Total Units	54 (-)	50 (-)	44 (-)	45 (-)
Total Assignments	2,051 (-)	1,297 (-)	1,437 (-)	1,518 (-)
Unique POs	12,944 (-)	8,618 (-)	8,468 (-)	9,832 (-)
N Obs.	11,993,940 (-)	3,894,885 (-)	3,728,734 (-)	4,370,321 (-)
C: Activity (per 100 Shifts)				
Street Stop	34.8 (112)	35.9 (111)	57.2 (148)	14.7 (60.9)
Traffic Stop	12.5 (56)	7.42 (41.9)	5.11 (36.8)	23.2 (75.4)
Arrest Index	1.52 (13.7)	1.61 (14)	1.58 (14.1)	1.38 (13)
Arrest NonIndex	10.4 (37.3)	13 (43.3)	10.8 (38.1)	7.73 (30.1)
Force	0.35 (6.07)	0.4 (6.51)	0.365 (6.19)	0.292 (5.53)
Complaints	0.17 (4.15)	0.251 (5.03)	0.201 (4.5)	0.0714 (2.7)

Note: Table presents summary statistics for officer-day level observations for days officer are present within the main analysis samples, with means and standard deviations (in parentheses). Geographic units refer to those which map directly to police districts. Assignment interacts the type of an assigned beat with watch and unit. The outcomes section takes the average across officer-day observations and multiplies them by 100 to get the mean per 100 officer-shifts. Panel A computes officer-level statistics then takes the observation-weighted average over officers.

Table 2: Summary Statistics of Teams Data

	Pooled	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)
Panel A: By Officer				
N. Teams (Obs.)	7899 (5884)	3787 (1605)	3890 (1815)	3858 (2609)
N. Unique Units	1.57 (0.904)	1.3 (0.61)	1.28 (0.601)	1.26 (0.57)
Panel B: Pooled				
Total Teams (District-Hour Obs.)	1836414 (0)	585304 (0)	539480 (0)	711630 (0)
Unique POs	11468 (0)	7186 (0)	7338 (0)	9029 (0)
Unique Units	25 (0)	25 (0)	22 (0)	22 (0)
Team Size (N. Active POs)	49.3 (18.2)	46.5 (17.7)	52.9 (19.1)	48.9 (17.4)
Panel C: Crimes Reported				
Index Violent	0.132 (0.381)	0.139 (0.394)	0.12 (0.362)	0.136 (0.384)
Murder	0.0026 (0.0532)	0.00214 (0.0484)	0.0023 (0.0494)	0.00321 (0.0594)
Crim. SA	0.00831 (0.0972)	0.00761 (0.0936)	0.0078 (0.0966)	0.00927 (0.101)
Robbery	0.0569 (0.246)	0.065 (0.263)	0.0531 (0.236)	0.0532 (0.238)
Agg. Assault	0.0643 (0.261)	0.0642 (0.262)	0.0567 (0.244)	0.0702 (0.271)
Index Property	0.491 (0.78)	0.563 (0.831)	0.466 (0.749)	0.45 (0.755)
Burglary	0.0863 (0.307)	0.12 (0.366)	0.0799 (0.292)	0.0634 (0.258)
Motor Vehicle Theft	0.0638 (0.259)	0.0849 (0.3)	0.0545 (0.239)	0.0534 (0.236)
Theft	0.339 (0.651)	0.356 (0.657)	0.329 (0.632)	0.331 (0.661)
Nonindex Drug	0.123 (0.411)	0.172 (0.476)	0.14 (0.444)	0.0693 (0.309)
Nonindex Simple Violent	0.302 (0.585)	0.321 (0.613)	0.293 (0.572)	0.295 (0.57)
Nonindex Weapon	0.0203 (0.146)	0.0176 (0.136)	0.016 (0.129)	0.0258 (0.164)
Nonindex All Other	0.419 (0.71)	0.437 (0.73)	0.415 (0.707)	0.407 (0.695)

Note: Table presents summary statistics of team / district-hour data for the main analysis samples, with means and standard deviations (in parentheses). Panel A computes officer-level statistics then takes the observation-weighted average over officers.

Table 3: Decomposition of Officer-Level Results

	Per 100 Shifts		Share of Var(Y) Explained				
	Mean(Y)	SD(PO)	Total	Var(PO)	Var(Assign)	2 x Cov(PO, Assign)	Cor(PO, Assign)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Street Stops							
2010-2012	36	30	0.15	0.072	0.063	0.013	0.097
2013-2015	57	43	0.17	0.085	0.069	0.017	0.11
2016-2019	15	15	0.15	0.061	0.067	0.021	0.16
Traffic Stops							
2010-2012	7.4	15	0.21	0.14	0.076	-0.003	-0.013
2013-2015	5.1	13	0.2	0.13	0.034	0.038	0.29
2016-2019	23	27	0.21	0.13	0.087	-0.003	-0.016
Arrests - Nonindex							
2010-2012	13	11	0.14	0.061	0.049	0.027	0.25
2013-2015	11	9.9	0.14	0.068	0.048	0.025	0.22
2016-2019	7.7	6.8	0.13	0.052	0.059	0.021	0.19
Arrests - Index							
2010-2012	1.6	1.2	0.021	0.007	0.017	-0.002	-0.11
2013-2015	1.6	1.4	0.025	0.01	0.013	0.001	0.048
2016-2019	1.4	1	0.019	0.006	0.013	-0.0001	-0.008
Uses of Force							
2010-2012	0.4	0.44	0.006	0.005	0.002	0.0001	0.024
2013-2015	0.36	0.4	0.005	0.004	0.001	-0.0002	-0.046
2016-2019	0.29	0.34	0.006	0.004	0.002	-0.0002	-0.036
Complaints							
2010-2012	0.25	0.22	0.003	0.002	0.001	0.00004	0.013
2013-2015	0.2	0.2	0.002	0.002	0.001	-0.0006	-0.21
2016-2019	0.071	0.056	0.002	0.0004	0.001	0.00004	0.029

Note: Table summarizes various statistics (columns) by outcome variable (panels) and sample period (rows). Estimates are bias-corrected using KSS. Columns (1) displays the mean of the outcome variable per 100 shifts, and (2) display the standard deviations of officer effects per 100 shifts recovered from estimating equation (1). Columns (3)-(6) display the components of variance decomposition divided by the variance of the outcome, with Column (3) being the summation of Columns (4)-(6). Column (7) is the correlation between officer and assignment effects.

Table 4: Cross-Outcome Correlation in Officer Effects

	Traffic Stops	Arrests - Nonindex	Arrests - Index	Use of Force	Complaints
	(1)	(2)	(3)	(4)	(5)
Street Stops					
2010-2012	0.111	0.347	0.344	0.168	-0.0297
2013-2015	0.0768	0.412	0.306	0.0984	-0.0336
2016-2019	0.483	0.453	0.409	0.277	0.162
Traffic Stops					
2010-2012		0.155	0.0539	0.0812	0.0786
2013-2015		0.147	0.0567	0.00167	0.0446
2016-2019		0.31	0.284	0.138	0.119
Arrests - Nonindex					
2010-2012			0.443	0.395	0.289
2013-2015			0.416	0.366	0.28
2016-2019			0.643	0.411	0.321
Arrests - Index					
2010-2012				0.324	0.11
2013-2015				0.341	0.0615
2016-2019				0.387	0.118
Use of Force					
2010-2012					0.536
2013-2015					0.634
2016-2019					0.604

Note: Table presents cross-outcome correlations between officer effects for each outcome and sample period. Estimates are bias corrected using KSS.

Table 5: Decomposition of Team Effects on Crime

	2010 - 2012			2013 - 2015			2016 - 2019		
	Mean(Y)	SD($\Delta\Omega$)	$\frac{Var(\Delta\Omega)}{Var(\Delta Y)}$ x 100	Mean(Y)	SD($\Delta\Omega$)	$\frac{Var(\Delta\Omega)}{Var(\Delta Y)}$ x 100	Mean(Y)	SD($\Delta\Omega$)	$\frac{Var(\Delta\Omega)}{Var(\Delta Y)}$ x 100
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Index Violent	0.139	0.022	0.167	0.120	0.020	0.172	0.136	0.023	0.198
Murder	0.002	0.003	0.177	0.002	0.002	0.102	0.003	0.003	0.127
Crim. SA	0.008	0.007	0.315	0.008	0.008	0.341	0.009	0.008	0.351
Robbery	0.065	0.012	0.105	0.053	0.012	0.129	0.053	0.012	0.141
Agg. Assault	0.064	0.014	0.144	0.057	0.011	0.098	0.070	0.015	0.150
Index Property	0.563	0.111	1.051	0.466	0.081	0.687	0.450	0.060	0.393
Burglary	0.120	0.032	0.409	0.080	0.022	0.302	0.063	0.016	0.187
Motor Vehicle Theft	0.085	0.017	0.170	0.055	0.015	0.201	0.053	0.014	0.177
Theft	0.356	0.080	0.867	0.329	0.064	0.602	0.331	0.049	0.349
Nonindex Drug	0.172	0.043	0.511	0.140	0.043	0.620	0.069	0.032	0.677
Nonindex Simple Violent	0.321	0.044	0.290	0.293	0.040	0.276	0.295	0.042	0.297
Nonindex Weapon	0.018	0.006	0.086	0.016	0.006	0.105	0.026	0.008	0.110
Nonindex All Other	0.437	0.101	1.098	0.415	0.105	1.247	0.407	0.098	1.092

Note: Table summary statistics by outcome variable / crime type (rows). Columns (1), (4), and (7) display the mean of the outcome variable at the district-hour level, and other columns display KSS bias corrected estimates for variance components of the first differences specification (equation (4)) across the three sample periods.

Table 6: Projection of Officer-Crime Effects onto Officer-Activity Effects (1 SD / Mean Y)

	Violent Crime			Property Crime			Drug Crime		
	2010-2012	2013-2015	2016-2019	2010-2012	2013-2015	2016-2019	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Street Stops	-0.315*	0.16	0.134	-0.139*	0.325**	-0.046	-0.002	0.26	-0.059
	(0.182)	(0.304)	(0.255)	(0.084)	(0.14)	(0.141)	(0.151)	(0.278)	(0.419)
Traffic Stops	-0.018	-0.2	-0.111	-0.062	0.175*	0.036	0.043	-0.061	0.173
	(0.145)	(0.185)	(0.222)	(0.078)	(0.097)	(0.129)	(0.112)	(0.149)	(0.413)
Arrests - Drug	0.288	-0.58*	0.086	-0.03	0.152	0.194**	0.396**	-0.08	0.822**
	(0.208)	(0.34)	(0.219)	(0.09)	(0.142)	(0.097)	(0.185)	(0.405)	(0.399)
Arrests - Non-Drug Nonindex	0.3	-0.306	-0.112	0.123	-0.298**	-0.103	0.177	-0.61**	0.162
	(0.219)	(0.291)	(0.28)	(0.106)	(0.135)	(0.148)	(0.179)	(0.267)	(0.428)
Arrests - Violent Index	0.013	0.089	0.13	-0.038	-0.095	-0.008	-0.084	-0.153	-0.289
	(0.157)	(0.194)	(0.17)	(0.074)	(0.092)	(0.083)	(0.124)	(0.184)	(0.276)
Arrests - Property Index	0.04	0.003	0.168	0.136	0.081	0.114	0.196	0.116	-0.252
	(0.197)	(0.205)	(0.207)	(0.11)	(0.12)	(0.133)	(0.152)	(0.169)	(0.259)
Uses of Force	0.465**	0.154	-0.129	0.002	-0.113	-0.148	0.03	0.672**	-0.229
	(0.212)	(0.28)	(0.201)	(0.09)	(0.119)	(0.096)	(0.173)	(0.305)	(0.362)
Complaints	-0.388	0.371	0.049	-0.135	-0.116	0.079	-0.164	0.369*	-0.004
	(0.264)	(0.244)	(0.148)	(0.107)	(0.116)	(0.072)	(0.196)	(0.218)	(0.206)
Mean Y	0.139	0.12	0.136	0.563	0.466	0.45	0.172	0.14	0.069
N. Obs	6933	7037	8583	6933	7037	8583	6933	7037	8583
R2	0.002	0.001	0.0002	0.0008	0.002	0.0005	0.002	0.002	0.0008

Note: Table displays the results of estimating equation (5): projecting officer effects on crime, recovered from estimating the first differences specification (equation (4)), onto shrunken officer effects on activities (recovered from estimating equation (1) but shrunken using the mixed models method). Coefficients are scaled to be the effect of 1 standard deviation increase in an officer-activity effect on officer-crime effects divided by the mean frequency of that crime (Mean Y). Mean Y refers to the mean of the initial outcome at the district-hour level prior to first differencing. P-values, represented by stars, are computed using KSS adjusted heteroskedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Effect of Team Average Activity Effects on Crime (1 SD / Mean Y)

	Violent Crime				Property Crime				Drug Crime			
	Pooled	2010-2012	2013-2015	2016-2019	Pooled	2010-2012	2013-2015	2016-2019	Pooled	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Street Stops	0.001	-0.004	0.002	0.011	-0.005	-0.002	-0.012*	-0.008	-0.004	-0.018*	0.024*	-0.039*
	(0.007)	(0.012)	(0.015)	(0.014)	(0.003)	(0.006)	(0.007)	(0.007)	(0.008)	(0.01)	(0.014)	(0.02)
Traffic Stops	-0.002	-0.009	0.008	-0.005	-0.002	0.004	0.004	-0.004	0.008	0.017**	0.001	0.008
	(0.007)	(0.009)	(0.011)	(0.014)	(0.004)	(0.005)	(0.007)	(0.008)	(0.006)	(0.007)	(0.011)	(0.02)
Arrests - Drug	0.001	-0.009	0.009	0.01	-0.005	-0.009	-0.003	-0.002	0.338***	0.265***	0.304***	0.612***
	(0.009)	(0.016)	(0.017)	(0.017)	(0.004)	(0.007)	(0.007)	(0.007)	(0.013)	(0.016)	(0.021)	(0.041)
Arrests - Non-Drug Nonindex	-0.015*	-0.01	-0.032*	-0.009	-0.01**	-0.007	-0.016*	-0.002	-0.053***	-0.045***	-0.063***	-0.058*
	(0.009)	(0.015)	(0.017)	(0.019)	(0.005)	(0.008)	(0.008)	(0.01)	(0.011)	(0.014)	(0.017)	(0.031)
Arrests - Violent Index	0.007	-0.003	0.012	0.017*	0.005	0.001	0.011**	0.002	-0.05***	-0.034***	-0.053***	-0.067***
	(0.006)	(0.012)	(0.011)	(0.01)	(0.003)	(0.005)	(0.005)	(0.005)	(0.008)	(0.011)	(0.013)	(0.016)
Arrests - Property Index	-0.007	-0.006	-0.004	-0.014	-0.003	-0.006	0.004	-0.006	-0.007	-0.021	0.009	-0.016
	(0.009)	(0.017)	(0.016)	(0.015)	(0.006)	(0.01)	(0.01)	(0.01)	(0.009)	(0.015)	(0.013)	(0.019)
Uses of Force	0.001	0.03*	-0.013	-0.025*	-0.004	0.007	-0.017**	-0.004	0.04***	0.036***	0.015	0.046*
	(0.009)	(0.016)	(0.017)	(0.014)	(0.004)	(0.007)	(0.007)	(0.007)	(0.01)	(0.014)	(0.016)	(0.026)
Complaints	0.003	-0.002	0	-0.001	-0.004	-0.011	0	-0.005	-0.003	0.001	0.011	0.004
	(0.009)	(0.017)	(0.015)	(0.009)	(0.004)	(0.007)	(0.007)	(0.005)	(0.011)	(0.015)	(0.015)	(0.014)
Mean DV	0.132	0.139	0.12	0.136	0.491	0.563	0.466	0.45	0.123	0.172	0.14	0.069
N. Obs	1836414	585304	539480	711630	1836414	585304	539480	711630	1836414	585304	539480	711630
Adj. R2	0.00002	0.00003	0.00004	0.00003	0.0002	0.00007	0.0003	0.0005	0.003	0.003	0.003	0.003

Note: Table displays results of estimating equation (6): regressing the number of reported crimes at the district-hour level on team composition as measured team averages of shrunken officer effects on activities (recovered from estimating equation (1) but shrunken using the mixed models method). District-Time controls used are District-Year-Month-Day of Week-Hour. Estimates are divided by the mean of the outcome variable and coefficients are scaled to be 1 standard deviation in the team average activity effect. Additional controls include polynomials of officer counts and team sizes, and control for exact-date fixed effects. Standard errors in parentheses are clustered at the District-Year-Month-Day of Week-Hour level. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Elasticities of Crime with Respect to Policing Activity

	Index Arrests	Nonindex Arrests	QOL Arrests	Street Stops	Traffic Stops
	(1)	(2)	(3)	(4)	(5)
Index Violent Crime					
Main	-0.028	-0.065	-0.039	-0.019	-0.013
	[-0.14, 0.084]	[-0.16, 0.027]	[-0.1, 0.022]	[-0.084, 0.046]	[-0.061, 0.035]
Raw FEs	0.0059	-0.088*	-0.076**	-0.022	-0.014
	[-0.11, 0.12]	[-0.19, 0.013]	[-0.15, -0.0049]	[-0.088, 0.044]	[-0.066, 0.038]
Predicted Teams	-0.067	-0.16	-0.12*	-0.04	0.017
	[-0.38, 0.24]	[-0.37, 0.052]	[-0.24, 0.012]	[-0.15, 0.066]	[-0.087, 0.12]
Main + FE Controls	-0.017	-0.11	-0.04	0.003	-0.006
	[-0.14, 0.1]	[-0.24, 0.021]	[-0.1, 0.024]	[-0.063, 0.069]	[-0.054, 0.042]
Index Property Crime					
Main	-0.02	-0.1***	-0.062***	-0.053***	-0.014
	[-0.086, 0.046]	[-0.15, -0.059]	[-0.092, -0.033]	[-0.086, -0.02]	[-0.038, 0.0097]
Raw FEs	-0.016	-0.085***	-0.056***	-0.052***	-0.016
	[-0.08, 0.049]	[-0.13, -0.037]	[-0.089, -0.023]	[-0.086, -0.018]	[-0.043, 0.011]
Predicted Teams	0.053	0.00069	0.0028	-0.027	-0.033
	[-0.11, 0.22]	[-0.092, 0.093]	[-0.055, 0.06]	[-0.079, 0.025]	[-0.082, 0.016]
Main + FE Controls	0.014	-0.073**	-0.04**	-0.026	-0.0059
	[-0.056, 0.084]	[-0.14, -0.0065]	[-0.071, -0.009]	[-0.059, 0.0073]	[-0.03, 0.018]

Note: Table displays the elasticities of index violent and property crime with respect to different policing activities (columns), with 95% confidence intervals displayed in brackets below, over the pooled sample (2010-2019). Elasticity estimates and confidence intervals are computed by estimating equation (7) regressing the number of crimes which occurred in a specific district-hour on the number of policing activities (e.g., street stops) and instrumenting for the policing activity using average team shrunken officer effects for that activity, then are divided by the mean of the outcome and multiplying by the mean of the activity. Rows within panels adjust the specification: 'Raw FEs' uses unshrunken (raw) officer effects to compute team averages for the first stage; 'Predicted Teams' uses the predicted rotations predict the officers working in that district-hour to construct the team average effects in the first stage; and 'Main + FE Controls' uses the main specification by adds controls in the second stage for team average officer effects on the other activities (e.g., controlling for team average effects for street stops, traffic stops, nonindex arrest, use of force, and complaints if the activity of interest is index arrests). QOL Arrests refer to 'quality of life' arrests, a subset of nonindex arrests. Officer effects are shrunken and computed separately over 2010-2012, 2013-2015, and 2016-2019 using equation (1). Standard errors used to compute confidence intervals are clustered at the year-month-day of week-district-hour level. *** p<0.01, ** p<0.05, * p<0.1.

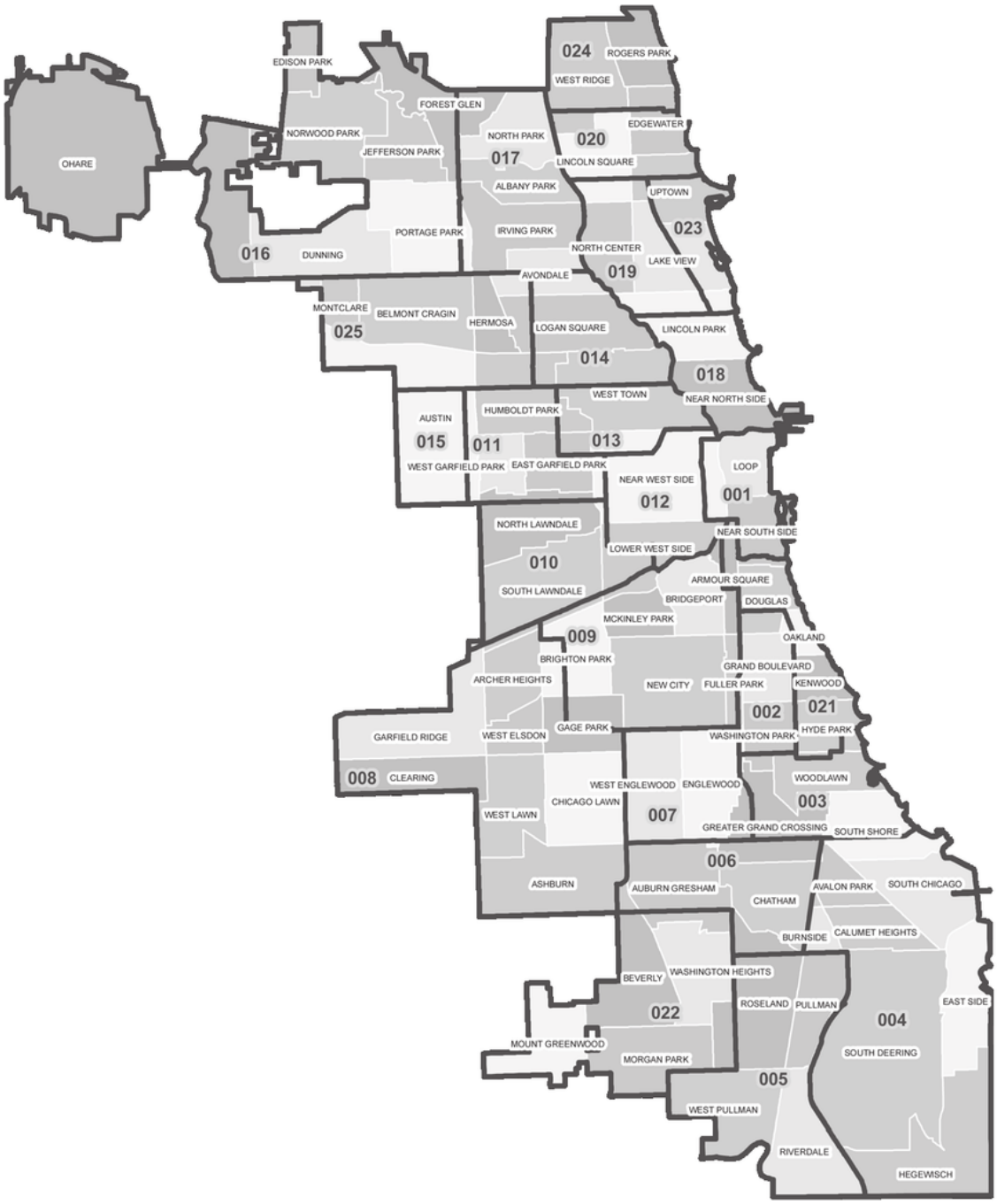
Table 9: Change in Outcomes and Costs from 1 SD Change in Single Officer Complaint and Force Effects over 100 Shifts

	1 SD Change in Complaint Effect			1 SD Change in Use of Force Effect		
	2010-2012	2013-2015	2016-2019	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)	(5)	(6)
A: Effect on Officer Activities						
Street Stops	1.59 (-)	3.49 (-)	-1.48 (-)	-4.88 (-)	-2.5 (-)	-4.35 (-)
Traffic Stops	-1.51 (-)	-0.68 (-)	-1.44 (-)	-1.33 (-)	0.02 (-)	-3.72 (-)
Arrests - Nonindex	-3.2 (-)	-2.55 (-)	-1.76 (-)	-4.1 (-)	-3.52 (-)	-2.99 (-)
Arrests - Index	-0.15 (-)	0 (-)	-0.12 (-)	-0.38 (-)	-0.51 (-)	-0.44 (-)
Uses of Force	-0.23 (-)	-0.25 (-)	-0.19 (-)	-0.44 (-)	-0.4 (-)	-0.34 (-)
Complaints	-0.22 (-)	-0.2 (-)	-0.06 (-)	-0.12 (-)	-0.13 (-)	-0.04 (-)
Total Cost ($1,000)	-75.95 (-)	-65.64 (-)	-35.64 (-)	-81.74 (-)	-73.36 (-)	-53.78 (-)
B: Effect on Crime Costs by Activity Effect ($1,000)						
Arrests - Index	-20.35 (29.2)	-0.93 (0.82)	-3.12 (27.22)	-51.44 (73.8)	121.12 (106.28)	-11.35 (99.12)
Arrests - Nonindex	59.06 (72.69)	70.33 (75.52)	13.8 (62.28)	75.77 (93.25)	96.81 (103.96)	23.43 (105.77)
Street Stops	-8.25 (9.74)	-20.48 (16.88)	3.26 (16.96)	25.29 (29.86)	14.68 (12.1)	9.6 (49.93)
Traffic Stops	10.52 (9.95)	6.75 (7.08)	2.93 (5.97)	9.27 (8.77)	-0.19 (0.2)	7.56 (15.41)
Arrests - Index (LM)	2.89 (4.3)	0.02 (0.18)	-1.37 (3.81)	7.32 (10.88)	-2.28 (23.23)	-5 (13.88)
Arrests - Nonindex (LM)	3.51 (9.43)	28.7*** (10.86)	-9.24 (7.74)	4.51 (12.1)	39.5*** (14.95)	-15.69 (13.14)
Street Stops (LM)	-4.81*** (1.32)	-4.01 (2.75)	-1.54 (2.44)	14.75*** (4.04)	2.87 (1.97)	-4.54 (7.18)
Traffic Stops (LM)	0.4 (1.27)	0.53 (1.28)	0.01 (0.97)	0.35 (1.12)	-0.02 (0.04)	0.02 (2.51)
C: Total Effect on Crime Costs from Activities ($1,000)						
Cost	40.97 (-)	55.67 (-)	16.87 (-)	58.88 (-)	232.42 (-)	29.24 (-)
Cost (LM)	1.99 (-)	25.23 (-)	-12.15 (-)	26.92 (-)	40.08 (-)	-25.21 (-)
D: Total Effect on Crime Costs from Team Avg. Effects ($1,000)						
Cost	-66.87 (-)	36.56 (-)	45.53 (-)	-94.28 (-)	95.95 (-)	57.24 (-)
Cost (LM)	-3.73 (-)	13.55 (-)	0.69 (-)	-2.2 (-)	14.21 (-)	4.79 (-)

Note: Table displays the effects of two counterfactuals: replacing a random officer with an officer with 1 SD lower complaint (use of force), in Columns (1)-(3) (Columns (4)-(6)) over 100 9-hour shifts. Panel A displays the expected change in other dimensions of the officer's activity based on the KSS bias corrected (co)variance estimates (Table 4), and the total social cost change summing over the change in activities times their associated costs, discussed in Section 8. Panel B displays the effects on crime costs, which combines estimates from regressing total crime costs at the district-hour level on officer activities (equation (7)) scaled by the associated changes in officer-level activities from Panel A. Panel C sums over the cost effects over each officer activity to recover a total change in crime costs. Panel D provides an alternative estimate for the numbers in Panel C by estimating the effect of team average officer effects on crime costs at the district-hour level (equation (6)) then multiplying the coefficients by the change in average team effects due to the 1 SD for a single officer in a team of the average size within each sample period, then summing over the scaled coefficients to recover a total cost change. Positive values denote cost increases, and negative values denote cost decreases (social savings). Standard errors, when displayed, are clustered at the year-month-day of week-district-hour level. *** p<0.01, ** p<0.05, * p<0.1.

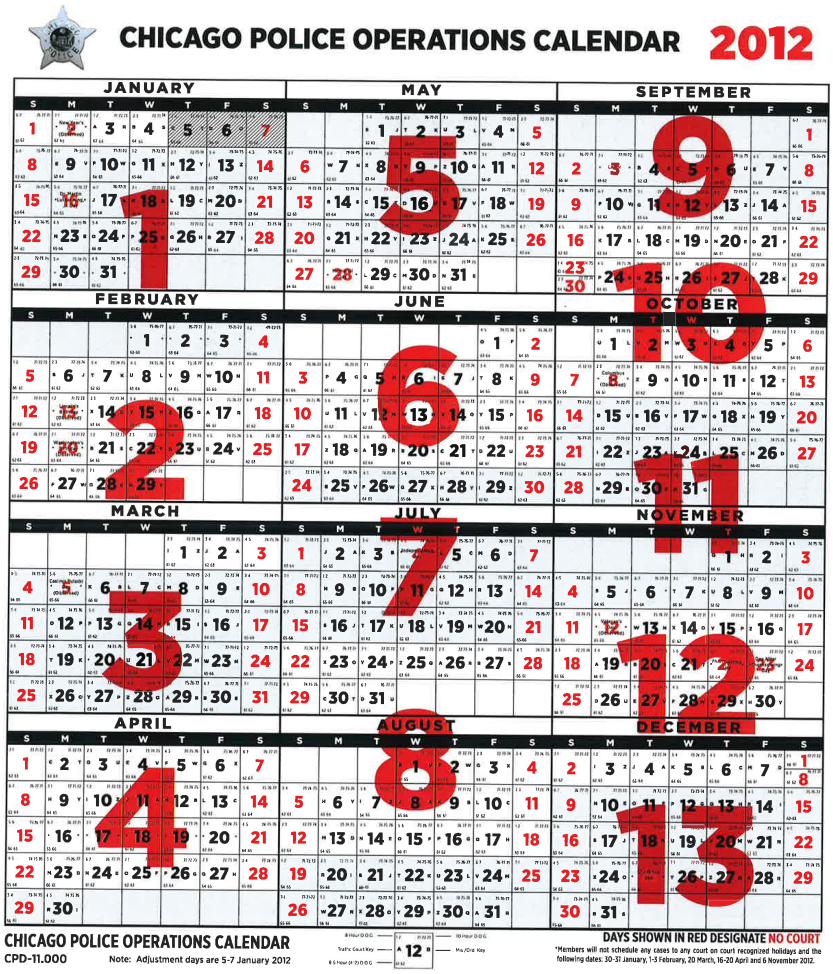
A Appendix A: Background and Data Figures and Tables

Figure A.1: Map of CPD Districts (2009) and Neighborhoods from Markovic (2009)



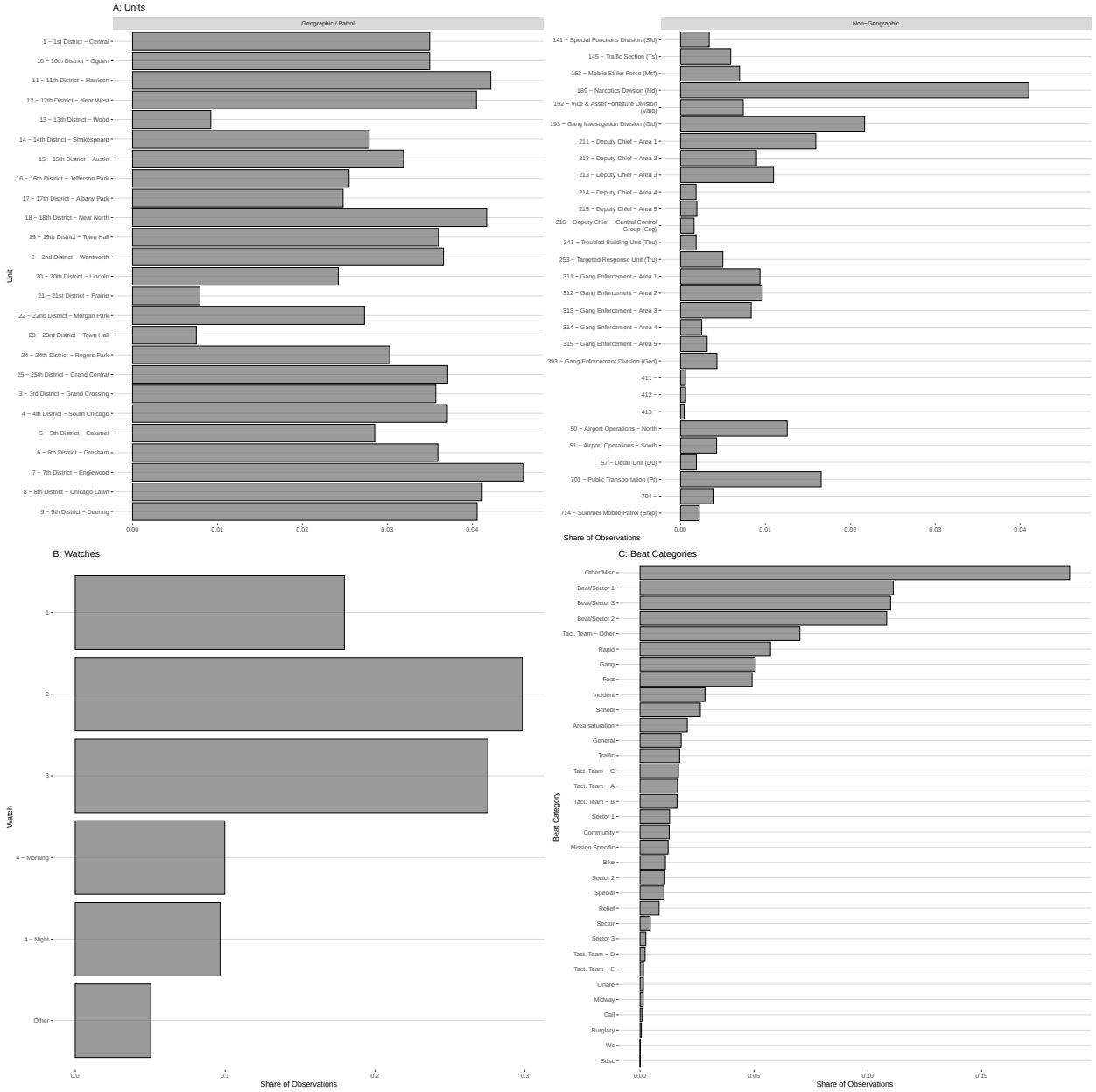
Note: Figure displays the Chicago Police Department districts numbered 1-25 from 2009 with labels for Chicago neighborhoods, sourced from Markovic (2009). In 2012, districts 13, 21, and 23 were absorbed by other districts mainly 12, 2, and 19, in the post-2012 district map.

Figure A.2: CPD Operations Calendar (2012)



Note: Figure displays an example of the CPD operations calendar for the year 2012.

Figure A.3: Share of Observations by Unit, Watch, and Beat Category



Note: Figure displays the share of observations in the analysis sample that corresponds to each unit (Panel A) separated by patrol units which map directly to police districts and non-geographic units which do not map to specific districts, each watch (Panel B) with watch 4 broken into morning and night corresponding to start times and all other watches beyond 1-4 pooled into "Other", and each beat category (Panel C). The interaction of unit, watch, and beat category produces a unique assignment or 'generalized assigned beat' (GAB).

Table A.1: Crimes and Arrests per District-Hour (2010 - 2019)

	Crimes	Uncleared	Cleared by Arrest	Clearance Rate
	(1)	(2)	(3)	(4)
Index Violent	0.132	0.110	0.022	0.169
Murder	0.003	0.002	0.001	0.377
Crim. SA	0.008	0.007	0.001	0.134
Robbery	0.057	0.052	0.005	0.094
Agg. Assault	0.064	0.049	0.015	0.231
Index Property	0.491	0.445	0.046	0.093
Burglary	0.086	0.082	0.005	0.053
Motor Vehicle Theft	0.064	0.060	0.004	0.066
Theft	0.339	0.302	0.037	0.108
Nonindex Drug	0.123	0.000	0.122	0.997
Nonindex Simple Violent	0.302	0.237	0.066	0.217
Nonindex Weapon	0.020	0.005	0.016	0.764
Nonindex All Other	0.419	0.324	0.095	0.227

Note: Table displays number of crimes, uncleared crimes, and crimes cleared by arrests at the district-hour level for Chicago from 2010 to 2019 along with the implied clearance rate by crime type.

B Appendix B: Detailed Discussion of Internal Validity

B.1 Switching and Connectivity

For officer-level outcomes, the identifying variation comes from officers working in different assignments across time. As Abowd et al. (2002) shows, officer and assignment effects can be identified for officers and assignments within a connected network, which depends on the structure of the bipartite graph produced by the connections between firms and workers. So, identification of officer and assignment effects requires that we observe officers moving between assignments with sufficient variety and frequency. Even within the connected set, if mobility (the degree of switching between officers and assignments) is low, then variance and covariance measurements from these identified effects may be biased due to weak connectivity or "limited mobility bias" (Andrews et al. (2008), Jochmans & Weidner (2019), Bonhomme et al. (2022)).[20]

As shown in Table 1, officers work in a relatively large number of assignments within a given year (over 4 assignments) but generally only work in one unit (1.17 per year). Across unit switches are rarer, with the average share of the most common assignment being 72% of an officer's observations within a year while the most common unit is 97% of observations in a given year.

Naturally, the best metric for the degree of connectivity is the size of the largest connected component (connected set) constructed from the network between officers and assignments. I follow the current literature pruning the network to find the largest connected component that is robust to the removal of any single officer, as in Kline et al. (2020) and Bonhomme et al. (2022). I will refer to this leave-one-out largest connected component as the "LLC" for brevity.

The frequency of switches between assignments results in almost all observations ($> 94\%$)

[20]Simple OLS regressions with officer and assignments fixed effects do not generally estimate the model. Effectively, because the matrix of officer and assignment fixed effects ($Z = [\theta', \psi']$) is sparse, in practices $Z'Z$ will often not have full rank.

belong to the LLC, as displayed in Table B.1. Similarly, the sample of only geographic units (1-25) and regular watches (1-4) accounts for around 76% of observations, while its LLC is still over 96% of the remaining sample. One concern may be that the high connectivity is only due to very short spells in different assignments and unit-watches which produce a connected network but the results are dependent upon officers switching due to short-term shocks. Table B.1 also shows the size of the sample and LLC share after filtering for observations with assignment spells at least 4 days and unit-watch spells at least 20 days, which correspond to $> 72\%$ and $> 86\%$ of the full sample, and their LLC's maintain over 95 of the resulting subsamples suggesting short-switches are not driving the connectivity. All LLC's are over 90% of their respective sample, with the exception of using unit-watch-beat assignment fixed effects, which results in only around 28% of the sample being kept in the LLC. On the other hand, the total samples for predicted on, high-outcome, and no partner samples are relatively small, generally around 35-50% of the complete sample being included — note that small samples for these in 2009 are due to a lack of stops data and a lack of predicted assignments for that year (they begin in 2010).

Table B.1: Samples and Connected Sets by Sample Periods

		2010-2012	2013-2015	2016-2019
		(1)	(2)	(3)
Total Counts	N. Shifts	4,050,005	3,885,244	4,610,868
	N. Unique POs	9,027	8,878	10,336
	N. Unique Assignments	1,642	1,792	1,966
Main	Share of Total	1	1	1
	LLC Share of Sample	0.96	0.96	0.95
Regular Units/Watches	Share of Total	0.77	0.77	0.76
	LLC Share of Sample	0.97	0.97	0.96
Long Unit-Watch	Share of Total	0.86	0.88	0.88
	LLC Share of Sample	0.97	0.97	0.96
Long GAB	Share of Total	0.73	0.73	0.73
	LLC Share of Sample	0.98	0.98	0.98
Donut	Share of Total	0.68	0.68	0.68
	LLC Share of Sample	0.98	0.98	0.98
High Outcome	Share of Total	0.57	0.65	0.4
	LLC Share of Sample	1	1	1
No Partners	Share of Total	0.38	0.38	0.39
	LLC Share of Sample	0.96	0.96	0.94
Predicted On	Share of Total	0.69	0.71	0.68
	LLC Share of Sample	0.97	0.97	0.96
Assign = Unit-Watch-Beat	Share of Total	1	1	1
	LLC Share of Sample	0.29	0.33	0.26
Assign = Unit-Watch	Share of Total	1	1	1
	LLC Share of Sample	1	1	1
Match Effects	Share of Total	1	1	1
	LLC Share of Sample	0.96	0.96	0.95

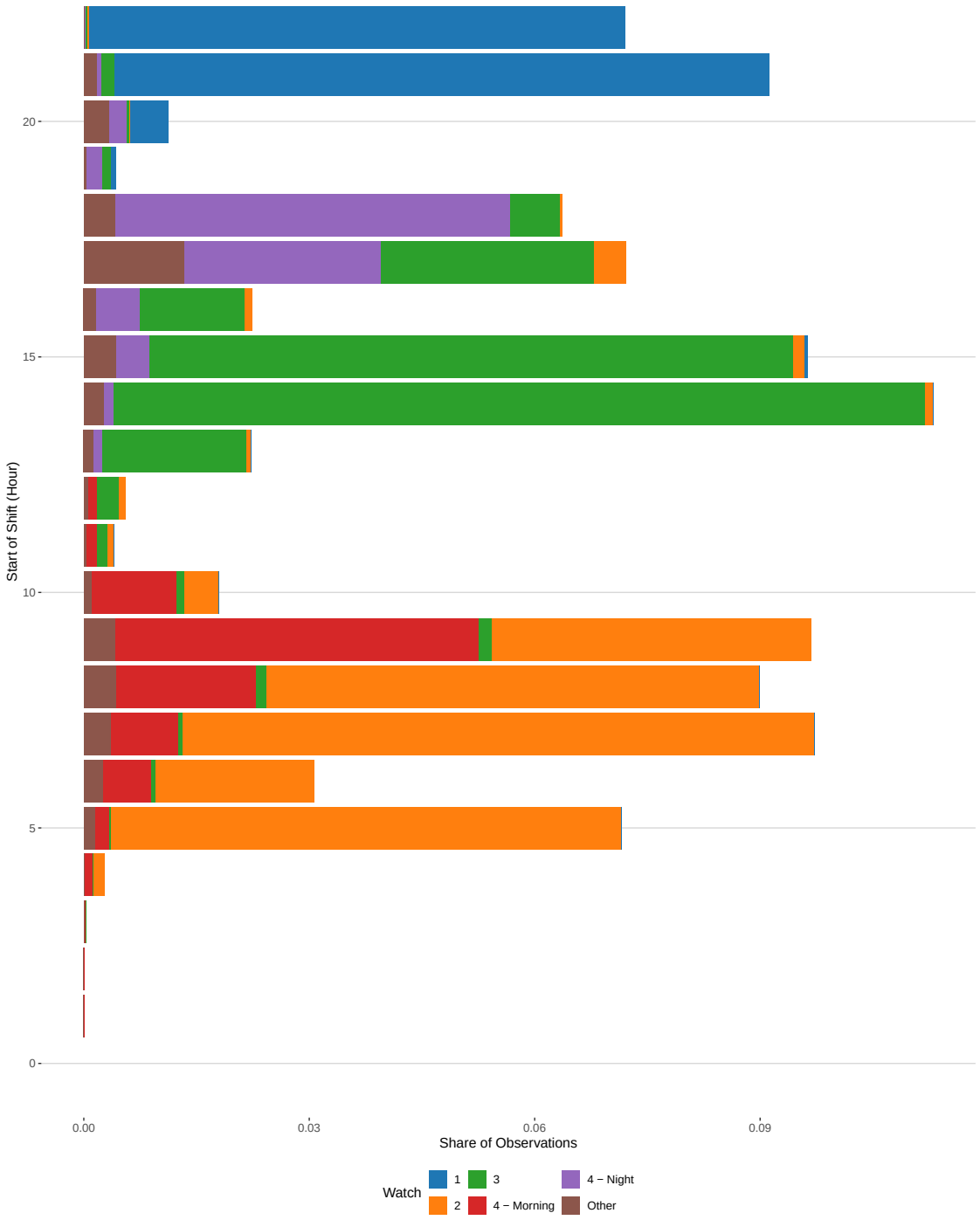
Note: Table presents summary of observations, total unique officers, and total unique assignments in each sample period, and the relative sizes of subsamples used along with the relative sizes of the largest leave-one-out connected component in terms of share of sample observations. Main refers to the main specification using the full samples within each period, along with robustness test samples described in Section 5.3.

For team-level outcomes, the variation comes from officers working in overlapping groups due to start and end times for the same watch not all being the same and ensuring teams

overlap. Figure B.1 displays the distribution of observed start times by watch number, which shows that each watch tends to have a 2-3 hour period in which almost all officers start their shifts, meaning there is overlap with officers working on the prior shift, producing variation in team composition. Additionally team size varies over the day. Figure B.3 displays the average team size by hour of day for teams classified into quantile based on their (within unit-year-quarter). Teams tend to be smallest in the early morning (around 5 am) and the differences between the groups is compressed. Teams tend to increase in size over the day and dispersion between groups grows, though there are small increases relative to the trends around 6 am and 2 pm, with teams tending to peak in size around 3pm and 11 pm — Figure B.4 displays the distributions of team sizes over each hour by team size group.[21]

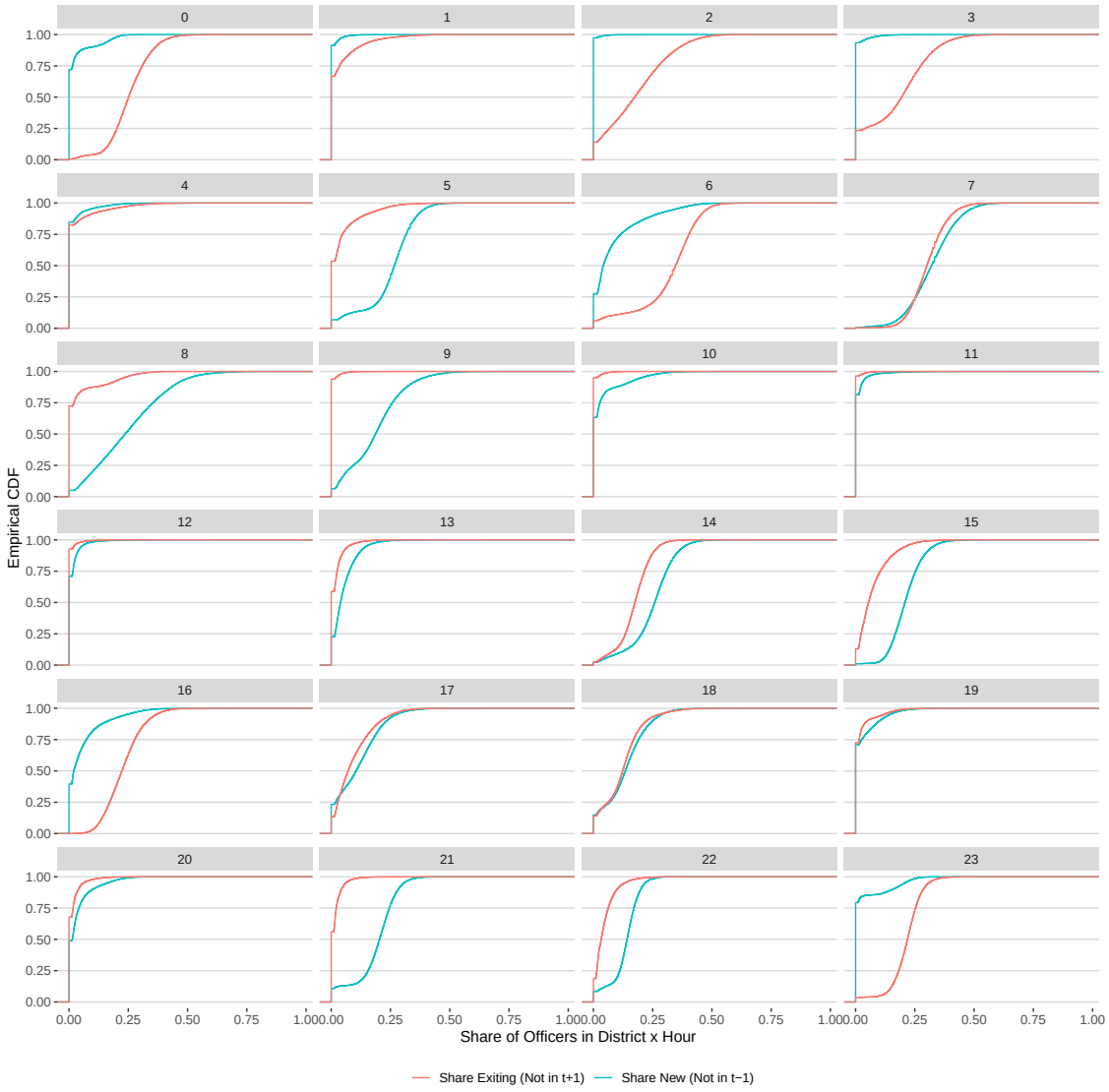
[21]Figure B.2 shows the empirical CDFs of the share of officers working in a district-hour that are new (were not working in hour $h-1$) and exiting (will not be working in hour $h+1$). In some times of day, there is not a significant change in the teams, such as 4-5am and 8 and 10 pm. However, some times tend to have many new entries, such as 7 am, 2 pm, and 9 pm, while others have many exiting officers such as 1-3 am and 3pm; some times have both many new entries and exits, such as 6 am, 5pm, and 11pm.

Figure B.1: Frequency of Shift Start Hours by Watch



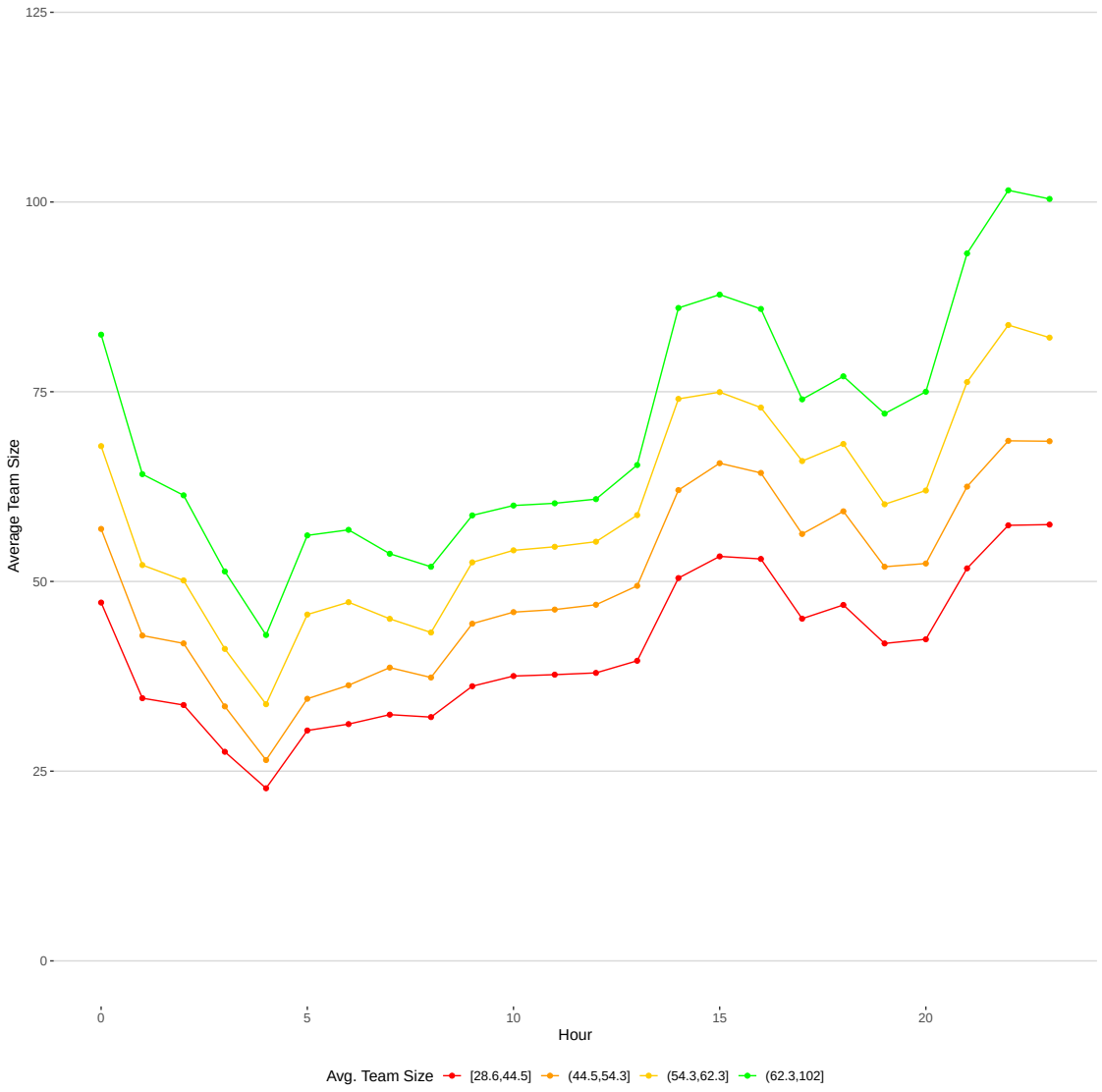
Note: Figure displays the share of observations (officer-shift) in the analysis sample by shift start hour, with colors indicating different watches, with watch 4 broken into morning and night corresponding to start times and all other watches beyond 1-4 pooled into "Other".

Figure B.2: ECDF of Share Entering and Exiting Officers by Team (District-Hour)



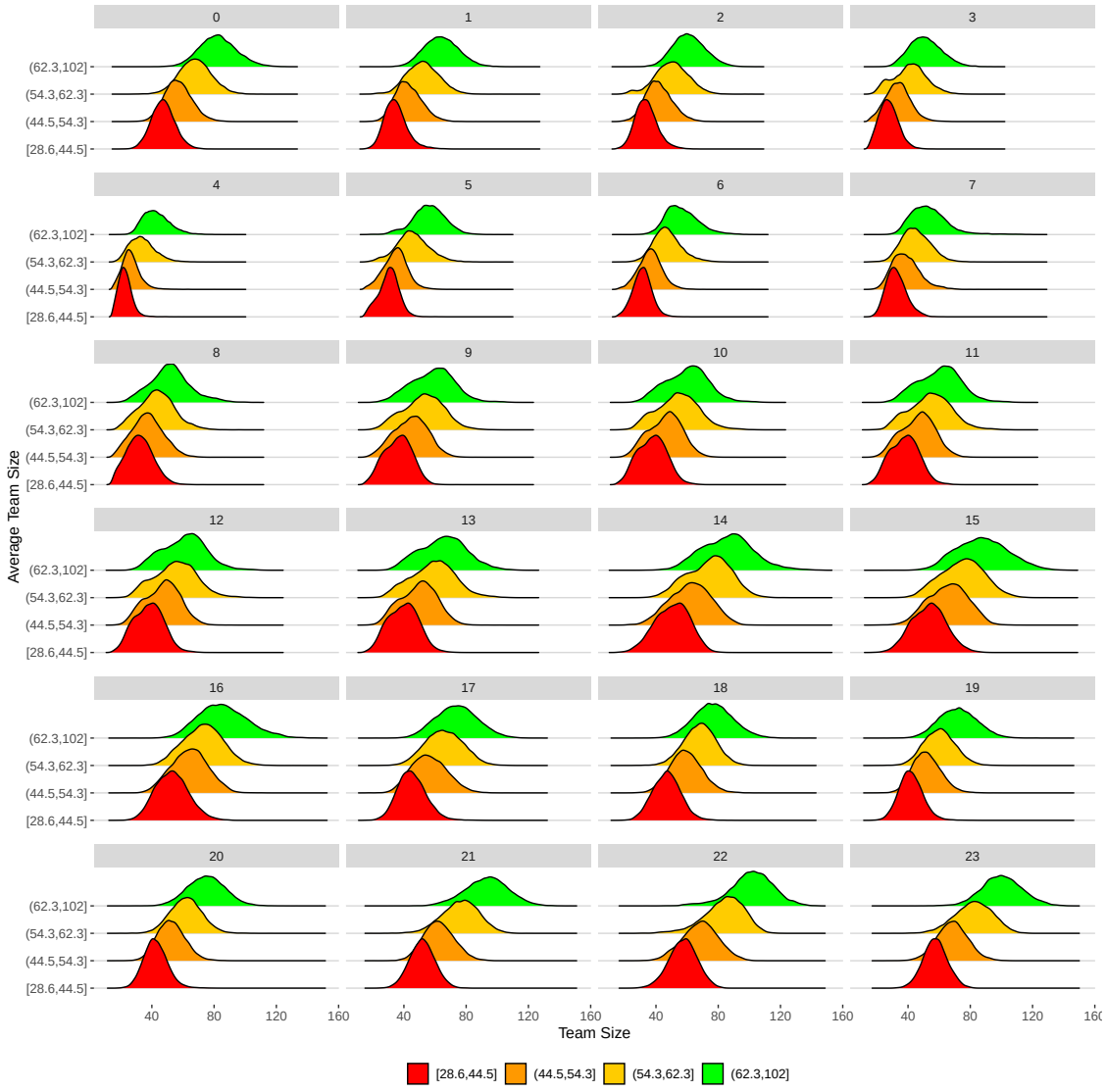
Note: Figure displays the empirical CDF of the shares of officers which are entering a team (started their shift in the hour) and share of officers which are exiting the team (ending their shift in the hour, meaning they will not be on the team in the next hour), with the denominator being the total number of officers on the team (district-hour) in the hour.

Figure B.3: Team Size by Hour in Day



Note: Figure displays the average team size (officers working in a district-hour) by hour of day with teams grouped into quantiles by their average size computed at the district-year-quarter level.

Figure B.4: Distribution of Team Sizes by Hour in Day



Note: Figure displays the distributions of team size (officers working in a district-hour) by hour of day with teams grouped into quantiles by their average size computed at the district-year-quarter level.

Similar to the officer-level model, identification of the officer effects in the team production model requires that the teams matrix (B) (along with the other fixed effects and controls) is non-singular. In the officer-assignment case the (largest) set of identified officer effects is in the largest connected component, in the team production model finding the largest identified set is more complicated. Bonhomme (2021) provides a pruning algorithm to iteratively remove indicators (officers) for whom effects cannot be identified, then removing any team

in which they are on, until convergence. To ensure non-singularity for all fixed effects, I prune the matrix containing officer indicators and district-time fixed effects to ensure both ω and γ are identified. The pruning process removes a small share of observations and officer/district-time indicators (see Table B.2), with at least 98% of observations remaining in the largest identified set, indicating that the degree of variation in team composition is sufficiently large.

Table B.2: Pruning and KSS Statistics for Crime Effects

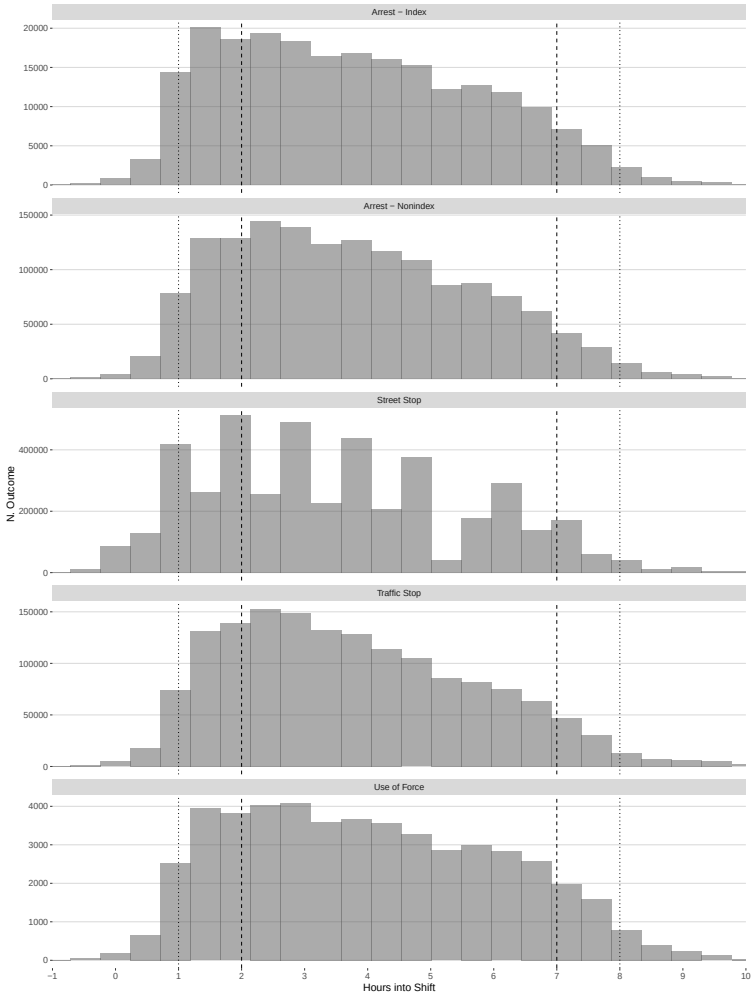
Spec Type	Date Controls	Sample Period	Initial Obs	Final Obs	Share Obs Remaining	Max PII
(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Differences		2010-2012	573156	573152	1.0000	0.9140
Pred. Team First Diff		2010-2012	548530	288448	0.5259	0.9854
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2010-2012	585304	584725	0.9990	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2010-2012	585304	585193	0.9998	0.2572
Officers (Main)	Year x Month x District x Watch x DoW	2010-2012	585304	585286	1.0000	0.2732
First Differences		2013-2015	516865	516832	0.9999	0.6048
Pred. Team First Diff		2013-2015	513332	236320	0.4604	0.9900
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2013-2015	539480	538719	0.9986	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2013-2015	539480	539316	0.9997	0.4256
Officers (Main)	Year x Month x District x Watch x DoW	2013-2015	539480	539316	0.9997	0.4241
First Differences		2016-2019	685627	685609	1.0000	0.8928
Pred. Team First Diff		2016-2019	672284	449616	0.6688	0.9900
Beat + Spec Officer	Year x Quarter x District x Watch x DoW	2016-2019	711630	711357	0.9996	0.9900
Officer + Sgt	Year x Quarter x District x Watch x DoW	2016-2019	711630	711519	0.9998	0.6303
Officers (Main)	Year x Month x District x Watch x DoW	2016-2019	711630	711519	0.9998	0.6363

Note: Table displays the results of the Bonhomme (2021) pruning algorithm on the crime data using the fixed effect specification with different date controls (year x district x day of week-8 hour time period (watch) x month or quarter) with observations are at the district-day-hour level or the first difference specification using actual or predicted (Pred.) team membership and observations at the first difference (within districts across hours) level. Max PII refers to the maximum leverage of any observation within the sample.

The identifying variation for team effects comes from officers entering and exiting their shifts at different times compared with other members of their current team. One concern is that officers may not be deployed at the beginning and end of their shifts meaning they are not an active component of their 'team' — a related concern is that officers strategically wait until the ends of their shifts to make more arrests to earn overtime. I do not find this to be

the case, Figure B.5 displays the distribution of activity metrics (stops, arrests, and use of force) over the duration into an officers' shift. It shows that officers rapidly increase their activity within 1 hour of their shift starting with activity levels peaking around 3 hours in, then slowly declining without any clear drop off right before their shift ends. Overall, this is inconsistent with officers being inactive at the starts or ends of their shifts and with officers strategically waiting for the end of their shifts to make arrests.

Figure B.5: Distribution of Outcomes by Number of Hours into Shift



Note: Figure displays the number of each outcomes which have time-stamp information (not complaints) grouped by the hour into the officer's shift they occurred – stops time stamps are at the hour-level pre-2016 producing more bunching at hours rather than half-hours. Most shifts are 9 hours long. Vertical lines denote 1 hour and 8 hours into the shift (dotted) and 2 hours and 7 hours into the shift (dashed).

Finally, while the canonical worker-firm model focuses on logged wages, making the

worker and firm effects identifying changes rather than levels, I use this model to study differences in levels, which vary widely. Part of this is due to the large number of zeros and the number of assignments and officers which never experience a specific outcome within a given time frame. I attempt to overcome this limitations in two ways. First, I estimate the model using the subset of assignments which frequently have a subset of outcomes occurring in at least 25% of shifts — this upper bound is to remove very uncommon assignments with very high means. In Table B.1, this high-outcome sample is between 35% to 65% of the data, due to changes in stopping frequency, and the LLCs are almost the entirety of the subsample always over 99%. Second, I estimate the model using a Poisson specification, which removes observations belonging to assignments which have no variation in the outcome variable.

B.2 Selecting Days-On

Given that there is sufficient variation to recover estimates for officer-level and team effects, the most direct threat to the identification of the parameters of interest in the following sections is that officers differentially select which days they work based on unobserved shocks to local crime rates or other conditions that allow for more enforcement, misconduct, or force. This would likely result in biased estimates of θ_i and ω_i because officers are choosing when to work based on the value of unobserved shocks ϵ. To overcome this, the exogenous variation for identification in the empirical sections will leverage the predetermined assignment of individual officers to work on specific days due to their DOGs being determined in the prior year through the bidding process.

The existence and importance of the operations calendar, in addition to CPD officers being governed by a strong union, makes such violations unlikely. As a first note, even after filtering uncommon beat types and allowing for occasional weekend rules and alternative rotations within some units (though still fixed based on rules), the 4-2 rotation is still, by far, the most common observed 'spell' type (meaning 4 days are observed worked in a row followed by 2 days not worked) at 43% of observations followed by 5-2 at 13%. All other spell

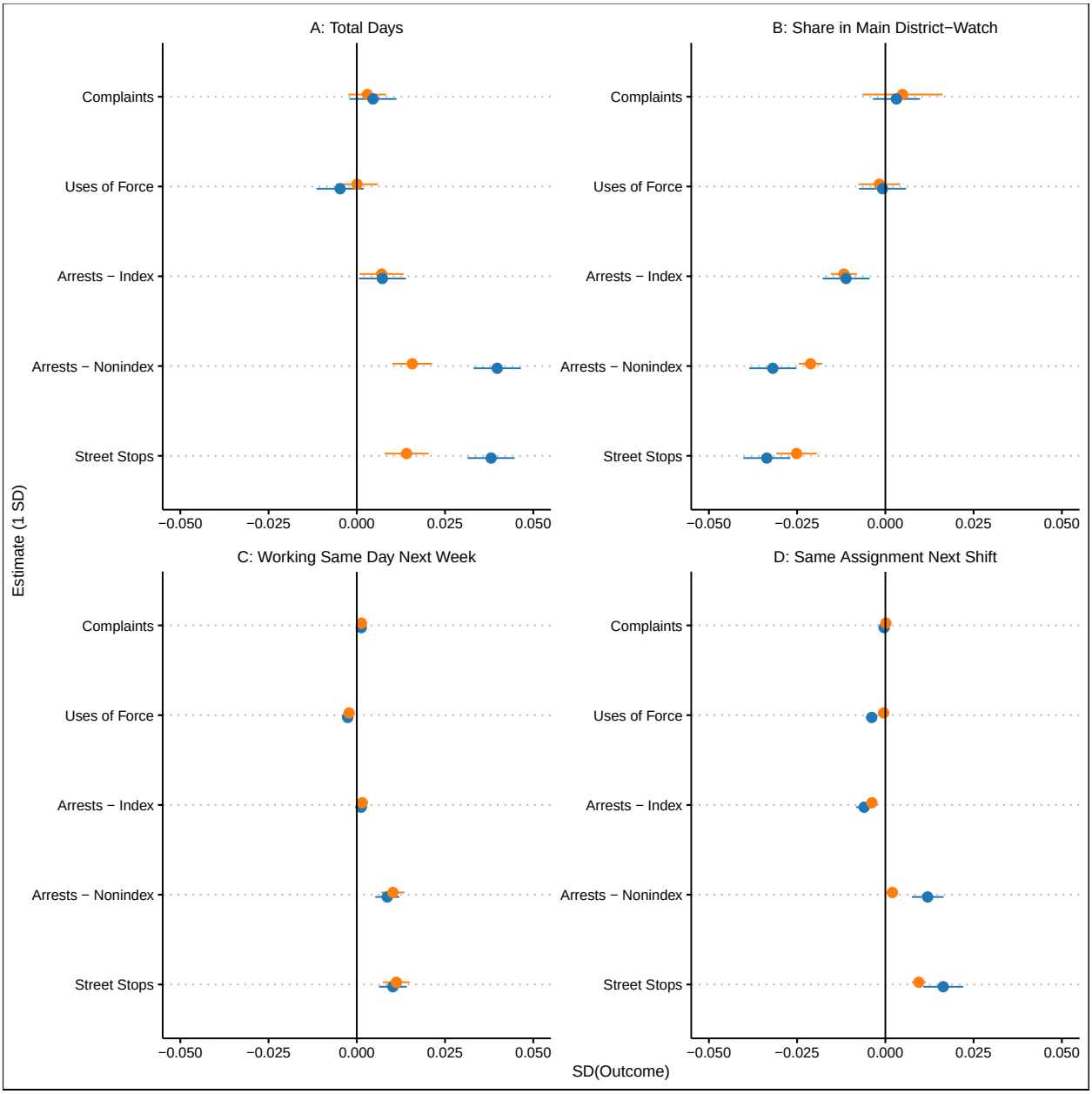
types are significantly less common (the next most common is 4-3 at 5%).

Beyond this, we can directly test the strength of the calendar following Gudgeon et al. (2023). Each year, I construct a predicted rotation for each officer based on the first set of observations, which follow a 4-2 (or, in a minority of cases, 5-2) rotation. Then I project this predicted rotation backward (if the first 4-2 is observed in the middle of January, for example) and forward throughout the rest of the year. This prediction is when the officer 'should' be working and 'should' be off based on a crude usage of the rotation system, ignoring all vacation days, injuries, training, potential for switching units, or changing DOGs. Nevertheless, as shown in Figure 1, which displays the share of correctly predicted days in each week of the year, this crude prediction is accurate at around 85% correctly predicted early on which decreases to 70% by the end of the year. For reference, if officers were randomly working 4 out of 6 days, the share of correct predictions would be around 45%. This consistency in predictive power of even a crude rotation is evidence of the binding nature of the calendar and the difficulty in officers deviating, even if they could observe shocks to local crime conditions on a day-to-day basis. As a robustness test, I re-estimate the main specifications for the officer-level analysis using only observations in which the officer is predicted based on the rotations calendar prediction to be working on that day, referred to as 'Predicted On'.

To test for systematic violations of the calendar, we perform the following tests. First, if officers cannot easily manipulate the calendar system, their outcomes should not relate to calendar-related statistics, such as the number of days they work in the year, the share of days they work in their main assignment, or their likelihood of working on the same date in the following week. As a note, around 90% of observations involve officers working in their most common (permanent) unit-watch in any given year, with the other 10% including in-year unit transfers and any deviations due to departmental demands. If, for example, high arrest officers preferred working more then we would expect average arrest rates to relate to the number of days they work or the amount of time they spend in other units. Similarly, if officers can manipulate their schedules then it is possible that outcomes on a specific date

would be predictive of deviations in the future or the probability of working the same date next week, both of which should be unaffected if the calendar binds. As shown in Panels A-C in Figure B.6, no outcome has an economically significant relationship with the share of days an officer spends in their main unit-watch, the total number of days they work, or their likelihood of working the same day next week with or without assignment (unit-year-watch) fixed effects — meaning a 1 SD increase in any activity has no more than a 0.05 SD increase or decrease in the outcome. Other work on the CPD have done alternative tests to verify the calendar system is obeyed in practice with positive results (e.g., Ba, Knox, et al. (2021), Ba, Bayer, et al. (2021), Gudgeon et al. (2023))— notably Gudgeon et al. (2023) finds that officers (the focus of this paper) do obey the calendar system and even lieutenants, to a lesser extent, who have much more discretion are still strongly influenced by it.

Figure B.6: Deviations from Permanent Unit, Calendar, and Assignment



Note: Figure displays the deviations from the operations calendar and steady assignment following the occurrence of a specific officer-level outcome (estimates scaled by 1 SD), and outcomes in the regressions are standardized. Points denote estimates with and without year-unit-watch fixed effects.

B.3 Endogenous Mobility

The next threat to identification is that errors are not conditionally independent or strictly exogenous with respect to the observables due to "endogenous mobility" (Card et al. (2013)). In this environment, officers are permitted to deliberately sort into specific assignments for any static reason. For example, they move from one assignment to another because of better working hours, lower crime rates, they will have more opportunities to commit misconduct, or because the commute from their home is shorter. These factors are fixed for the assignment, and thus are captured in ψ_J, and any match-specific component (e.g., commuting distance) can be assumed to be mean zero, and thus unobserved shocks to outcomes are not an issue.

However, Card et al. (2013) outlines three sources of endogenous mobility which can also arise in the policing environment, and 'event study' figures are traditionally used to assess the prevalence of such issues. Figure 2 displays such event studies for all main outcomes, with officers working in assignments classified by the frequency of the respective outcome within that year, then the assignments are binned into 4 groups based on their quantile of their average outcome (Q1-4, where 4 is the highest). Figure 2 displays the event study plots for each core outcome, showing the average outcome by switch type (e.g., Q4 to Q1, Q1 to Q3, etc.) for the six shifts before and after an officer changes their assignment (note that many shifts occur after less than 6 days, but these contribute to the means as well).

First, officers may sort into assignments due to match-specific effects, meaning that $\epsilon^k_{i,t}$ contains a component an officer-assignment component which officers are using to determine where to work. If this match component is not mean zero, this is a violation of strict exogeneity; then when officers move between assignments we would expect that the officers who move from assignments with lower outcomes to ones with higher outcomes would experience out-sized gains in their individual frequency of that outcome compared with officers moving from higher to lower. For example, if match effects exist for arrests, then officers who are moving from relatively low arrest assignments to higher arrest ones should see a much larger gains in their individual arrests compared with officers moving from high arrest assignments to low arrest

ones. Figure 2 shows that this is not the case. Officers who move from Q1 assignments to higher quantiles do not experience out-sized increases in outcomes compared to the declines of those moving from Q4 assignments to lower ones.

Second, there may be a drift component in $\epsilon_{i,t}^{k}$ if officers learn during their assignments and this is predictive of future movements. While we expect officers to learn as they work, this would be an issue if officers who are experiencing upward (or downward) trajectories in their outcomes systematically switch to higher (or lower) outcome assignments because this drift component would then produce upward bias in the variation attributed to assignment effects because higher outcome assignments would receive officers who are on (unobservably) upward trends and the opposite for lower outcome assignments. As in Figure 2 there tend to be downward trends for Q4 officers prior to switches, however this is true for switches to all quantiles from Q4, meaning that the downward trend is not indicative of a downward move. Furthermore, the decline for Q4 and increase for Q1 (when present) is driven by the day right before the move and there is a symmetric pattern after they move.

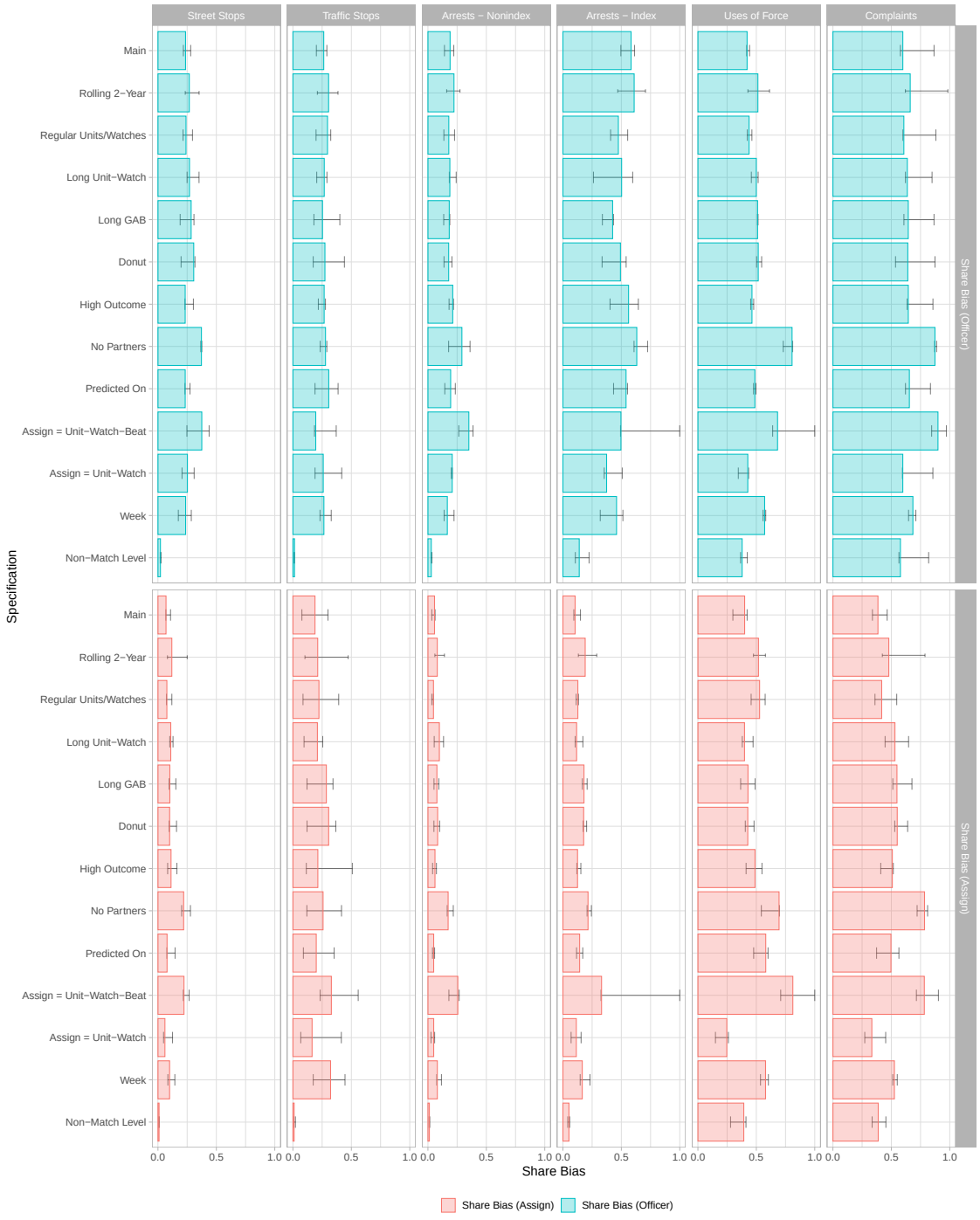
Third, if officers switch assignments based on transitory shocks to their outcomes — for example, if an officer who receives a complaint due to a random shock is then moved to a higher complaint assignment the next day this would produce biased estimates. However, shown in the event study plots, there is no sudden jump or drop in average outcomes among switchers, and this is bolstered by the results in Panel D in Figure B.6, which displays the relationship between officer outcomes and the likelihood of working in the same assignment in the following shift. As shown, a specific outcome on a given day has virtually zero relationship both with and without unit-year-watch fixed effects.

While these patterns are generally consistent with the exogenous mobility assumptions, I probe the robustness of my main results using alternative samples which should alleviate any remaining concerns. I use alternative samples for the main analysis focusing on long spells within assignments (at least 4 shifts), long spells within unit-watches (at least 20 shifts), a 'donut' approach where I remove the first and last observation within an officer's spell for

their consecutive shifts in an assignment (unit-watch-beat type), and I aggregate data at the week level and the officer's assignment within a week is their modal assignment. These alternative specifications will avoid over-weighting short and sporadic spells which may be more susceptible to endogenous mobility concerns and counting the dips/rises in outcomes immediately before and after a switch if present. Additionally, if there are concerns over over- or under- specificity in definition of an assignment, I redefine it as either Unit-Watch or Unit-Watch-Beat as well. Finally, Appendix D.2 discusses how much variation is explained when using a full match-effects specification.

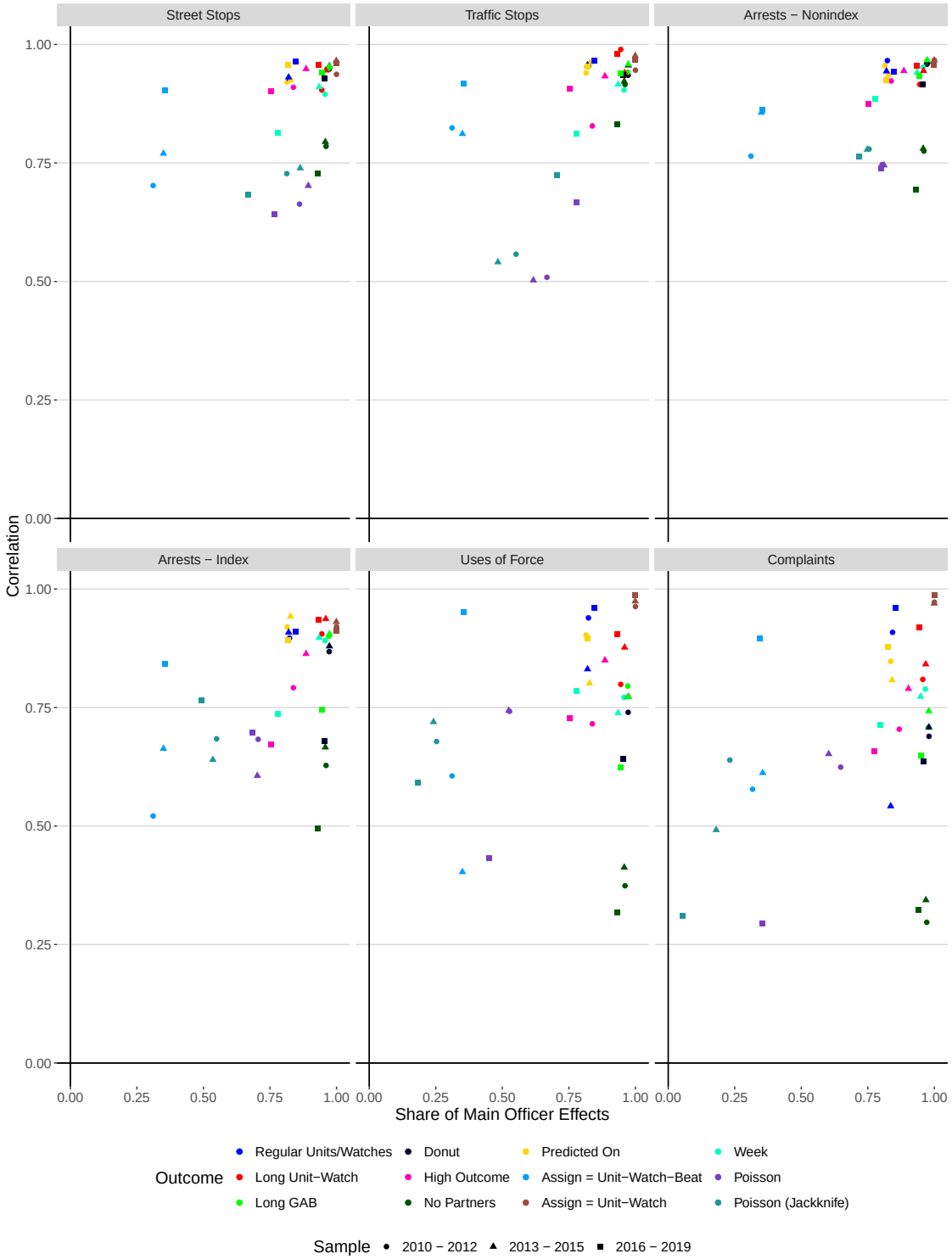
C Appendix C - Additional Figures and Tables for Results

Figure C.1: Share of Plug-In Variance of Officer and Assignment Effects Attributable to Bias



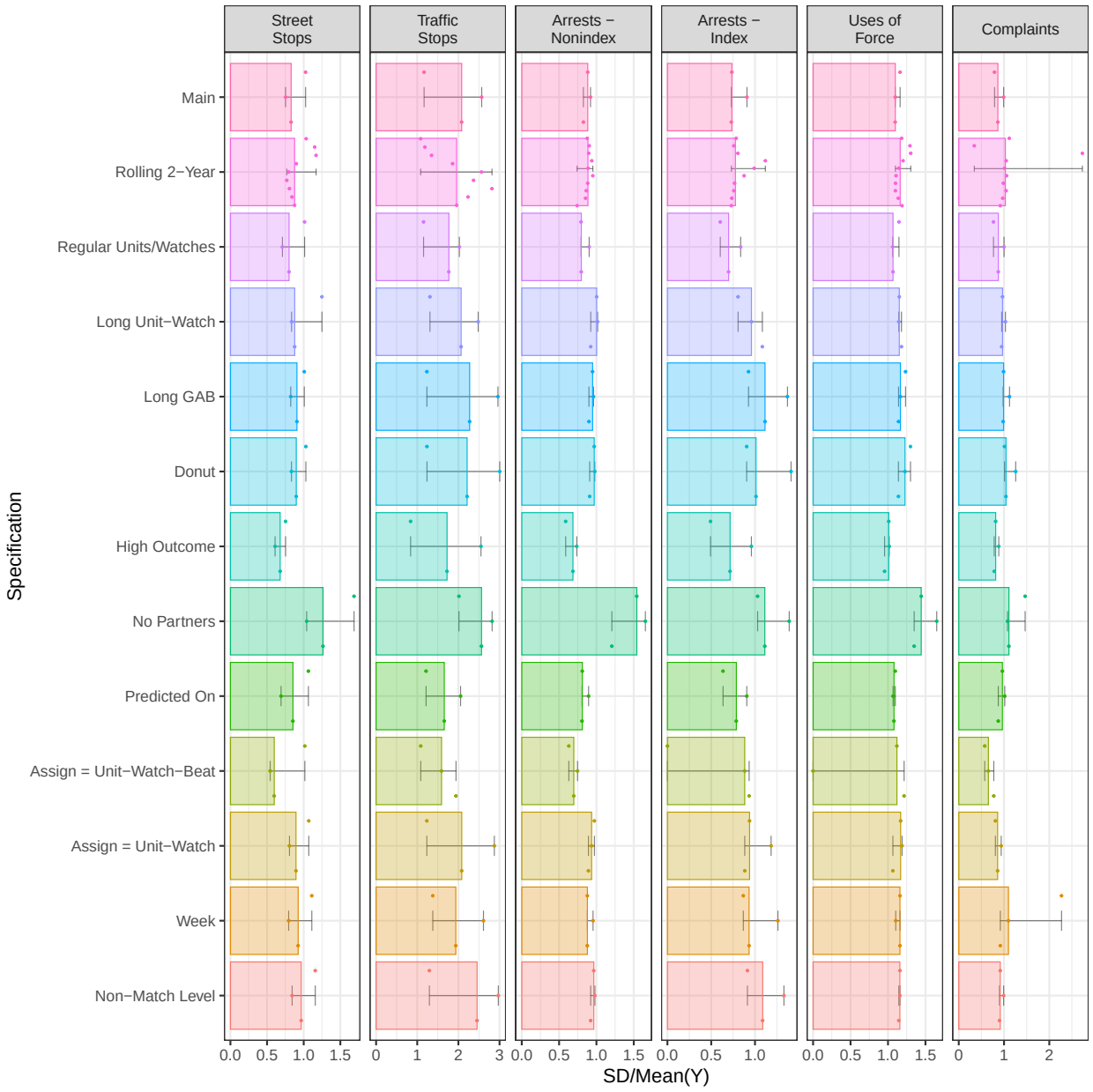
Note: Figure displays the share of the plug-in (non-bias corrected) variance in officer and assignment effects which are attributed to bias based on the KSS bias correction across specifications. The bars denote the median share of bias and the error bars denote the minimum and maximum share of bias across samples within the specification.

Figure C.2: Correlation between Officer Effects Across Specifications



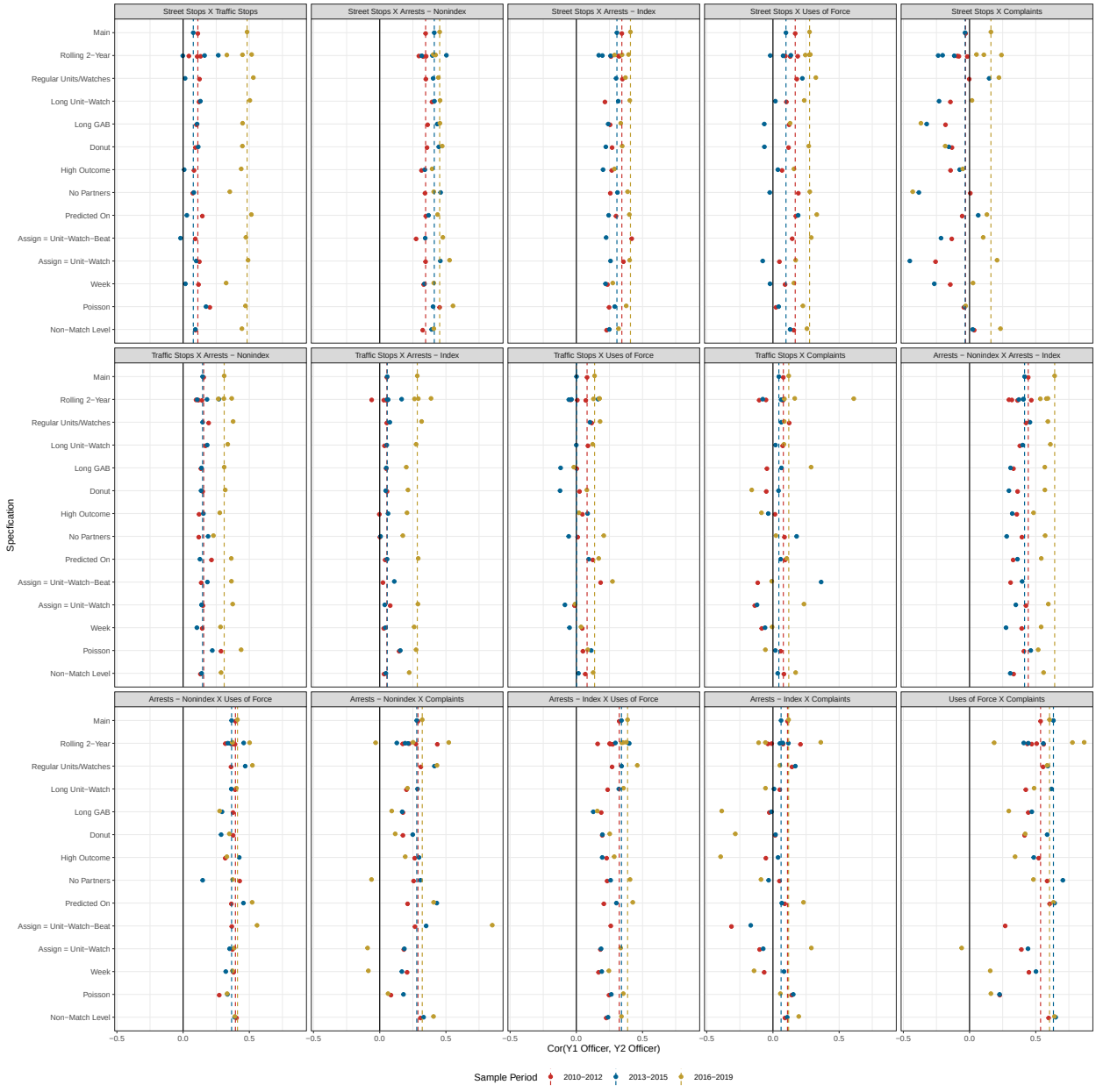
Note: Figure displays the (non-bias corrected) correlation (y-axis) between officer effects recovered from the main specification with officer effects recovered in other specifications over the same two year sample period. The x-axis contains the share of officer effects recovered in the main specification that were also recovered in the alternative specification.

Figure C.3: Magnitude of Officer and Assignment Effect Heterogeneity by Specification



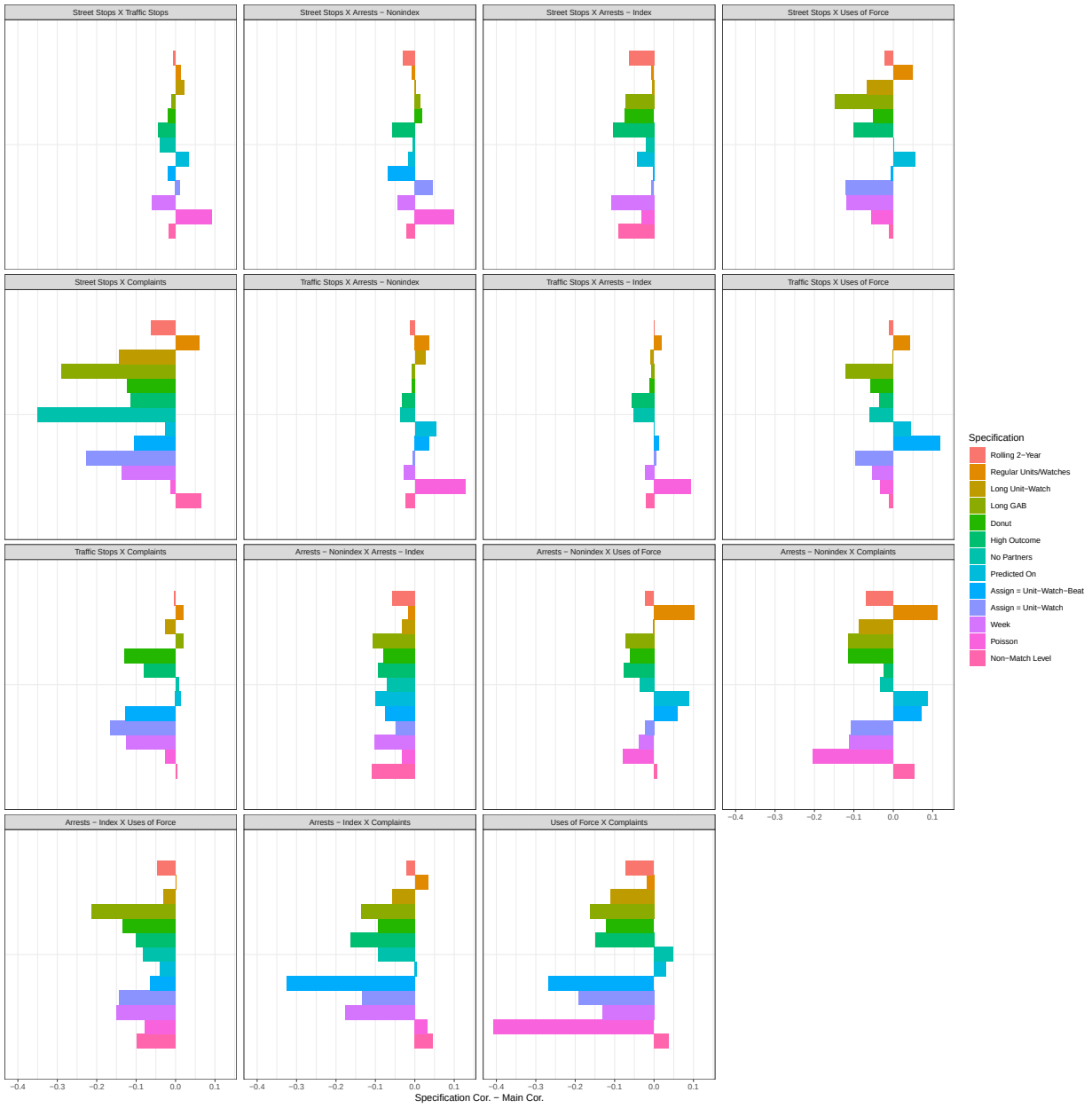
Note: Figure displays the KSS bias-corrected standard deviations of officer and assignment effects (recovered from equation (1)) across different specifications scaled by the mean of the outcome, with each point denoting an estimate for a specific time period, the bar denoting the median across these estimates, and the confidence interval denoting the minimum and maximum estimates. Main refers to the main specification using the 2010-2012, 2013-2015, 2016-2019 sample periods; Rolling 2-Year refers to using 10 2-year rolling sample periods (2009-2010, 2010-2011,..., 2018-2019); Regular Unit/Watches refers to the main specification but subsetting to geographic units only and watches 1-3; Long Unit-Watch (GAB) refers to selecting for spells in which the officer spends at least 20 (4) consecutive days in the unit-watch (assignment, "GAB"); Donut refers to removing the first and last day of each officer-assignment spell; High Outcome selects for assignments which have average outcomes (all main outcomes summed) of at least 0.25 within a year; No Partners removes observations in which the officer has another officer in their car; Predicted On includes observations where the officer was predicted to be working on that exact day, watch, and unit; Assign = Unit-Watch-Beat (Unit-Watch) replaces the assignment definition in the main specification, Unit-Watch-Beat Category, with Unit-Watch-Exact Beat (Unit-Watch); Week refers to aggregating the data to the weekly level; Non-Match Level computes leverages at the observation, rather than officer-assignment match, level.

Figure C.4: Cross-Outcome Correlations in Officer Effects across Specifications



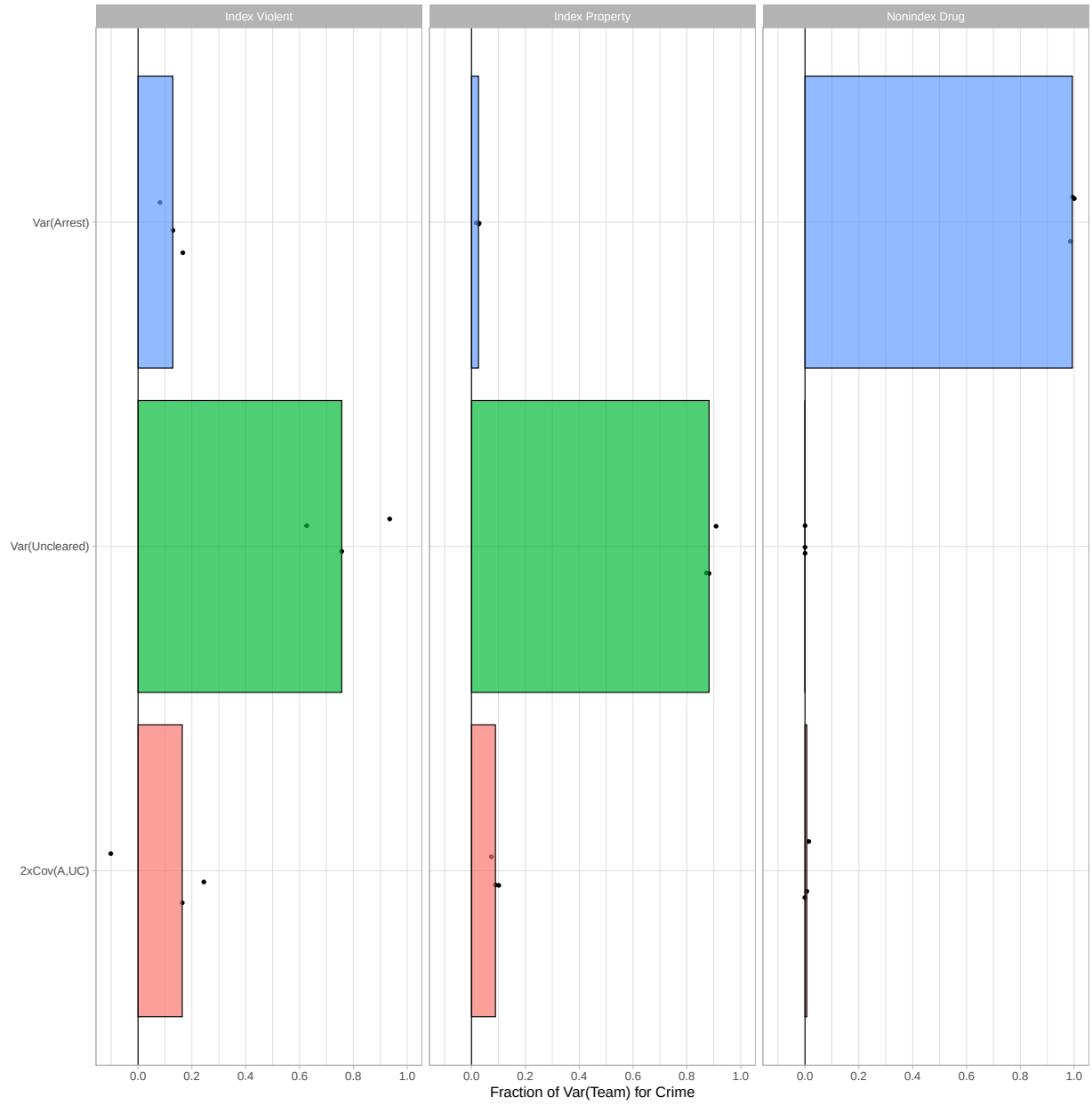
Note: Figure displays the KSS bias-corrected (excluding Poisson) cross-outcome correlations for officer effects across different specifications, with each point denoting an estimate for a specific time period. Dashed lines correspond to the value of the main specification. Main refers to the main specification using the 2010-2012, 2013-2015, 2016-2019 sample periods. See Section 5.3 for descriptions of all specifications.

Figure C.5: Difference between Cross-Outcome Correlations in Officer Effects across Specifications and Main



Note: Figure displays the difference between cross-outcome correlations of alternative robustness specifications and that of the main specification, with the alternative specifications described in Section 5.3 — all are KSS bias corrected except for the Poisson. Bars denote medians across sample periods.

Figure C.6: Decomposing Team Effects on Crime into Arrests and Uncleared Crimes



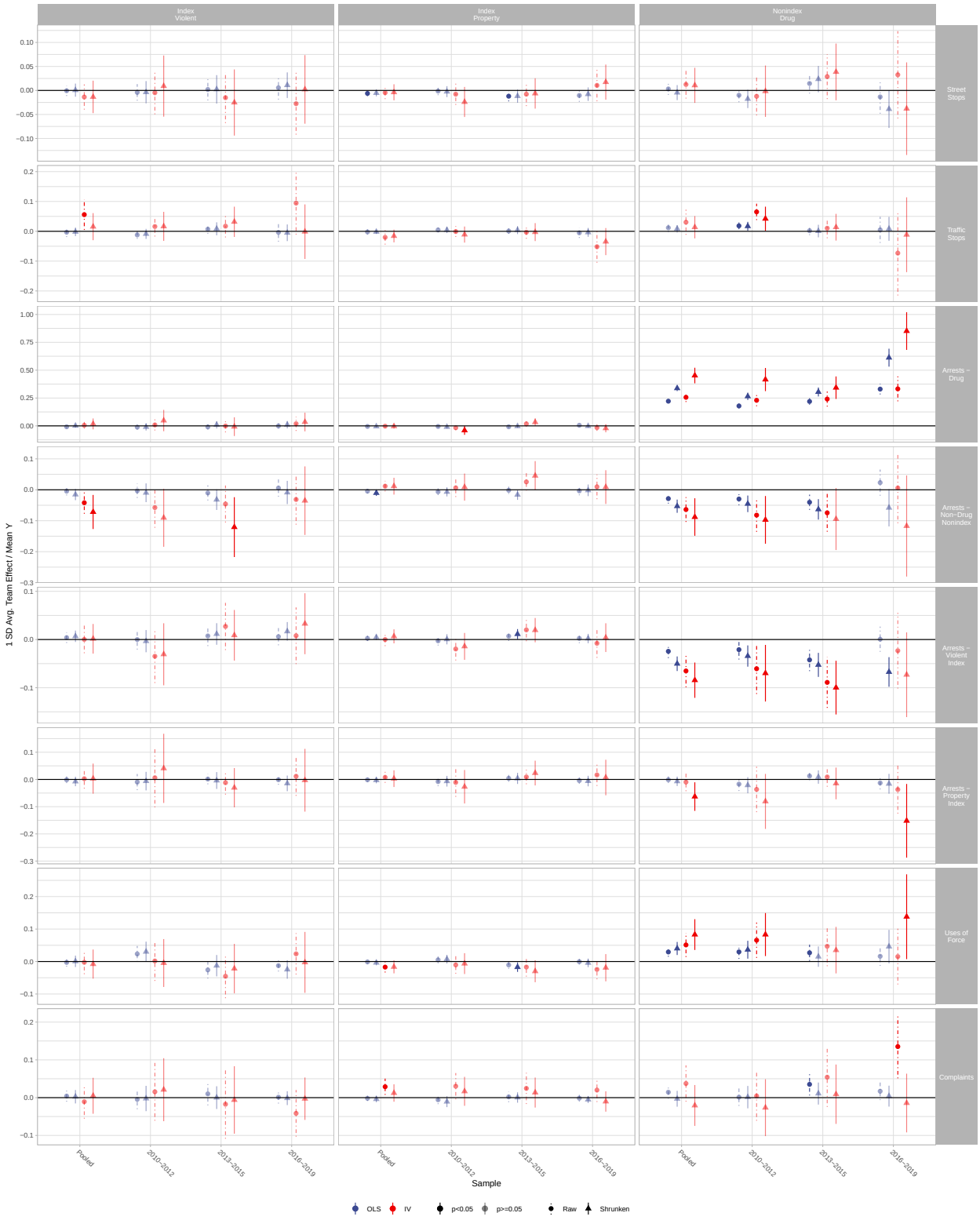
Note: Figure displays the KSS bias corrected variance and covariances for team effects (Ω) for the components of total crime (arrests and uncleared crimes) scaled by the overall variance in team effects on crime. Estimates of team effects recovered from estimating the main first differences specification on 3 multi-year samples, and the bars denoting the median of these estimates. Variance components can be negative due to bias correction.

Figure C.7: Influence of Teams on 911 Calls



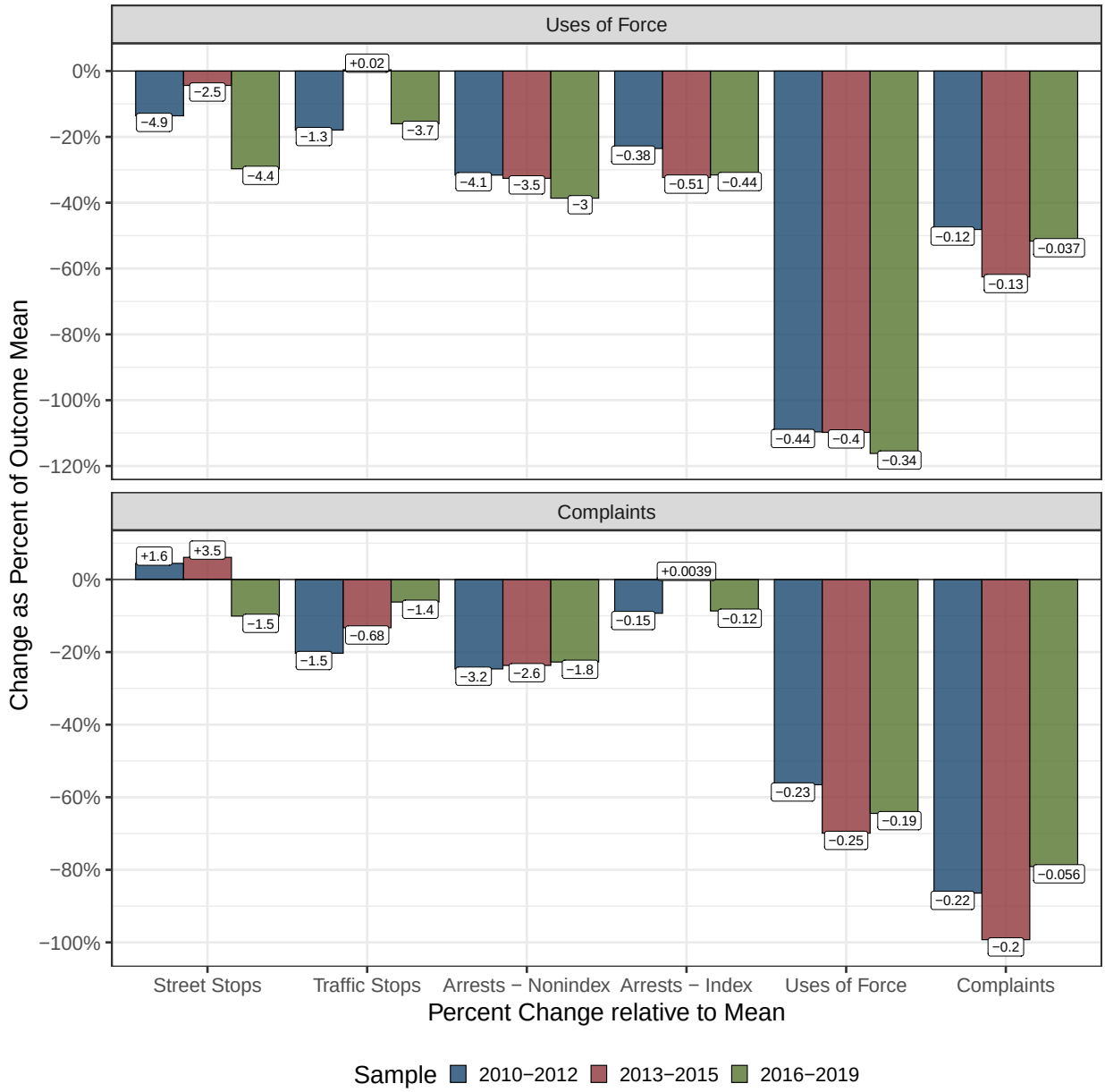
Note: Figure displays the KSS bias corrected standard deviations (left panel) and variance (right panel) of team effects (Ω) for 911 call outcomes (y-axis), scaled by the mean of the outcome and by the overall variance of the outcome, respectively. Team effects were estimated using the fixed effect, equation (3), and first difference specification, equation (6), and each point corresponds to one of three multi-year samples under different fixed effects or using actual or predicted team composition, while the bars denote the median of these estimates, and the colors correspond to whether or not the calls included in the outcome included reporting calls. Zero or negative values reflects plug-in estimates being entirely attributable to bias.

Figure C.8: Effect of Team Average Predicted Officer Effects on Crime



Note: Figure displays coefficients and associated 95% confidence intervals for the effect of 1 standard deviation increase in team average officer propensity for each officer-level activity (rows) on different crime types (columns), from estimating equation (6) for OLS and IV refers to instrumenting for team average officer propensities with predicted team average officer effects based on the rotation calendar, over the pooled sample (2010-2019) or one of the three sample periods. Officer effects are either not shrunken (raw) or shrunken. Standard errors are clustered at the year-month-district-day of week-1 hour level.

Figure C.9: Effect of 1 SD Decrease in Use of Force and Complaint Effects



Note: Figure displays the effect of a 1 SD decrease in either use of force or complaint effects (top and bottom panels) on officer effects for each outcome across sample periods (bars), with the y-axis being scaled as a percent of the mean of the outcome and the labels denoting the change in terms of outcomes per 100 shifts. The results are based on the (co)variance relationships implied by Tables 4 and 5. For example, in the 2010-2012 sample, a 1 SD decrease in use of force propensity corresponds to around a 16% decline in street stops relative to the mean or around 4.9 fewer street stops over 100 shifts.

Table C.1: Cross-Crime Correlations in Officer Effects - Part 1

	Reported Crimes								
	Murder	Rape	Robbery	Agg. Assault	Property	Burglary	MV Theft	Theft	Drug
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Reported Crimes									
Violent	0.243	0.386	0.736	0.754	0.258	0.11	0.242	0.251	0.153
	[0.17, 0.31]	[0.2, 0.49]	[0.62, 0.77]	[0.73, 0.81]	[0.2, 0.26]	[0.079, 0.16]	[0.15, 0.26]	[0.12, 0.28]	[0.086, 0.22]
Murder		0.0498	0.0346	0.0962	0.061	0.0808	0.026	0.0659	0.0369
		[-0.01, 0.17]	[-0.07, 0.14]	[0.083, 0.22]	[-0.061, 0.15]	[-0.14, 0.095]	[-0.14, 0.12]	[-0.039, 0.14]	[-0.01, 0.11]
Crim. SA			-0.0134	0.14	0.529	0.233	0.344	0.534	0.0225
			[-0.11, 0.064]	[-0.22, 0.16]	[0.26, 0.56]	[0.23, 0.26]	[0.31, 0.52]	[0.17, 0.55]	[-0.042, 0.026]
Robbery				0.258	0.0224	-0.0464	0.0428	-0.0101	0.134
				[0.22, 0.34]	[-0.069, 0.068]	[-0.092, 0.026]	[-0.13, 0.096]	[-0.056, 0.13]	[0.034, 0.31]
Agg. Assault					0.156	0.0741	0.156	0.11	0.11
					[0.047, 0.16]	[0.055, 0.13]	[0.063, 0.2]	[0.02, 0.15]	[0.073, 0.12]
Property						0.637	0.49	0.946	0.077
						[0.56, 0.68]	[0.47, 0.75]	[0.92, 0.96]	[0.027, 0.14]
Burglary							0.336	0.409	0.0357
							[0.23, 0.37]	[0.27, 0.47]	[-0.02, 0.078]
MV Theft								0.285	0.0791
								[0.22, 0.7]	[0.035, 0.11]
Theft									0.0789
									[0.029, 0.12]
Drug									
S. Violent									
Weapon									
All Other									
911 Calls									
Domestic									
Assault									
Robbery									
MV Theft									

Note: Table presents cross-outcome correlations for team membership effects on crime and 911 calls, using the first difference specification in equation (6). The first row for each statistic refers to the median estimate across three sample periods and pooled across specifications, and the second row displays the minimum and maximum values.

Table C.2: Cross-Crime Correlations in Officer Effects - Part 2

	Reported Crimes			911 Calls				
	S. Violent	Weapon	All Other	Domestic	Assault	Robbery	MV Theft	Theft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reported Crimes								
Violent	0.414	0.236	0.37	0.368	0.399	0.362	0.128	0.178
	[0.34, 0.49]	[0.074, 0.31]	[0.24, 0.44]	[0.3, 0.59]	[0.25, 0.56]	[0.3, 0.37]	[-0.022, 0.15]	[0.11, 0.31]
Murder	0.0342	0.0486	0.096	0.0463	0.0912	0.0281	-0.0657	0.136
	[0.033, 0.17]	[-0.089, 0.05]	[-0.038, 0.1]	[0.032, 0.15]	[0.014, 0.22]	[-0.015, 0.11]	[-0.2, 0.008]	[0.11, 0.16]
Crim. SA	0.0881	0.0763	0.608	0.128	0.0145	-0.0448	-0.00707	-0.0681
	[0.019, 0.12]	[-0.075, 0.17]	[0.55, 0.68]	[0.12, 0.18]	[-0.032, 0.2]	[-0.055, 0.001]	[-0.078, 0.032]	[-0.09, 0.21]
Robbery	0.309	0.246	0.118	0.346	0.134	0.474	0.0955	0.0846
	[0.21, 0.56]	[-0.13, 0.33]	[0.02, 0.21]	[0.02, 0.41]	[0.025, 0.59]	[0.43, 0.59]	[0.003, 0.36]	[0.055, 0.43]
Agg. Assault	0.257	0.159	0.132	0.348	0.391	0.105	-0.0483	0.142
	[0.21, 0.43]	[0.078, 0.26]	[0.066, 0.14]	[0.23, 0.48]	[0.35, 0.44]	[0.079, 0.21]	[-0.094, 0.28]	[0.065, 0.15]
Property	0.287	0.132	0.774	0.202	0.227	0.0247	0.0353	0.214
	[0.22, 0.29]	[0.062, 0.27]	[0.6, 0.83]	[0.17, 0.35]	[0.22, 0.3]	[-0.056, 0.1]	[-0.008, 0.063]	[0.17, 0.27]
Burglary	0.208	0.0989	0.388	0.113	0.159	-0.0234	-0.0754	0.0643
	[0.13, 0.25]	[-0.023, 0.24]	[0.34, 0.44]	[0.059, 0.24]	[0.037, 0.18]	[-0.15, 0.008]	[-0.18, -0.02]	[-0.071, 0.12]
MV Theft	0.134	0.03	0.429	0.136	0.0894	0.0112	0.196	-0.0217
	[0.086, 0.16]	[-0.12, 0.35]	[0.28, 0.64]	[0.12, 0.31]	[-0.018, 0.15]	[-0.12, 0.18]	[0.11, 0.3]	[-0.19, -0.018]
Theft	0.25	0.133	0.732	0.188	0.258	0.0199	0.0314	0.249
	[0.2, 0.28]	[0.12, 0.2]	[0.55, 0.86]	[0.17, 0.29]	[0.23, 0.32]	[-0.036, 0.14]	[0.002, 0.041]	[0.23, 0.36]
Drug	0.166	0.285	0.208	0.133	0.167	0.0873	-0.016	0.0438
	[0.07, 0.38]	[0.17, 0.39]	[0.037, 0.24]	[0.12, 0.34]	[0.11, 0.32]	[-0.052, 0.29]	[-0.086, 0.017]	[0.03, 0.11]
S. Violent		0.086	0.307	0.613	0.568	0.267	0.218	0.3
		[0.044, 0.39]	[0.23, 0.32]	[0.51, 0.69]	[0.56, 0.64]	[0.11, 0.41]	[0.13, 0.27]	[0.1, 0.5]
Weapon			0.116	0.275	0.115	0.306	-0.114	-0.12
			[0.052, 0.33]	[0.27, 0.28]	[0.099, 0.15]	[0.17, 0.34]	[-0.2, 0.33]	[-0.27, 0.18]
All Other				0.363	0.281	0.0355	0.0383	0.117
				[0.27, 0.44]	[0.21, 0.33]	[0.033, 0.14]	[-0.054, 0.063]	[0.054, 0.24]
911 Calls								
Domestic					0.495	0.236	0.158	0.204
					[0.45, 0.61]	[0.067, 0.44]	[-0.089, 0.16]	[0.18, 0.37]
Assault						0.123	0.0933	0.313
						[0.073, 0.56]	[0.071, 0.21]	[0.13, 0.35]
Robbery							0.136	0.0156
							[-0.069, 0.33]	[-0.25, 0.27]
MV Theft								0.109
								[-0.24, 0.17]

Note: Table presents cross-outcome correlations for team membership effects on crime and 911 calls, using the first difference specification in equation (6). The first row for each statistic refers to the median estimate across three sample periods and pooled across specifications, and the second row displays the minimum and maximum values.

Table C.3: Decomposition of Team Effects on Crime across Specification

	Mean(Y)	SD(Team)/Mean(Y)			Var(Team)/Var(Y) x 100		
	FE	FE	FD	PFD	FE	FD	PFD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Index Violent	0.14	0.4	0.17	0.14	2.1	0.17	0.11
	[0.12, 0.14]	[0.37, 0.65]	[0.16, 0.17]	[0.13, 0.18]	[1.5, 5.3]	[0.17, 0.2]	[0.11, 0.22]
Murder	0.002	3.6	0.96	0.56	3.7	0.13	0.034
	[0.002, 0.003]	[0, 6]	[0.93, 1.3]	[0.35, 0.81]	[-3.9, 7.8]	[0.1, 0.18]	[0.02, 0.07]
Crim. SA	0.008	0	0.93	0.6	-2	0.34	0.16
	[0.008, 0.009]	[0, 2.9]	[0.89, 0.96]	[0.56, 0.75]	[-2.2, 5.7]	[0.31, 0.35]	[0.11, 0.19]
Robbery	0.053	0.086	0.22	0.11	0.037	0.13	0.034
	[0.05, 0.07]	[0, 0.63]	[0.18, 0.23]	[0.08, 0.22]	[-1.6, 2.5]	[0.11, 0.14]	[0.02, 0.15]
Agg. Assault	0.064	0.16	0.21	0.18	0.17	0.14	0.11
	[0.06, 0.07]	[0, 0.51]	[0.19, 0.21]	[0.17, 0.18]	[-0.59, 1.5]	[0.1, 0.15]	[0.09, 0.11]
Index Property	0.47	0.5	0.17	0.15	9.6	0.69	0.52
	[0.45, 0.56]	[0.35, 0.52]	[0.13, 0.2]	[0.14, 0.17]	[5.8, 9.7]	[0.39, 1.1]	[0.46, 0.8]
Burglary	0.08	0.18	0.26	0.18	0.24	0.3	0.12
	[0.06, 0.12]	[0, 0.87]	[0.25, 0.28]	[0.16, 0.2]	[-0.53, 8]	[0.19, 0.41]	[0.08, 0.23]
Motor Vehicle Theft	0.055	0.81	0.26	0.21	4.7	0.18	0.11
	[0.05, 0.08]	[0.76, 1.2]	[0.2, 0.27]	[0.15, 0.26]	[3.3, 8]	[0.17, 0.2]	[0.06, 0.27]
Theft	0.33	0.63	0.19	0.17	11	0.6	0.45
	[0.33, 0.36]	[0.46, 0.67]	[0.15, 0.22]	[0.17, 0.19]	[6.4, 12]	[0.35, 0.87]	[0.44, 0.59]
Nonindex Drug	0.14	1.2	0.3	0.55	14	0.62	1.3
	[0.07, 0.17]	[1.2, 1.7]	[0.25, 0.46]	[0.4, 0.64]	[14, 19]	[0.51, 0.68]	[1.3, 1.7]
Nonindex Simple Violent	0.29	0.48	0.14	0.12	6.4	0.29	0.22
	[0.29, 0.32]	[0.41, 0.51]	[0.14, 0.14]	[0.11, 0.12]	[4.4, 7]	[0.28, 0.3]	[0.17, 0.24]
Nonindex Weapon	0.018	1.2	0.31	0.3	3.5	0.11	0.071
	[0.02, 0.03]	[0, 2.1]	[0.29, 0.36]	[0.25, 0.34]	[-4.6, 7.6]	[0.09, 0.11]	[0.06, 0.15]
Nonindex All Other	0.41	0.44	0.24	0.22	6.8	1.1	0.94
	[0.41, 0.44]	[0.27, 0.49]	[0.23, 0.25]	[0.22, 0.24]	[2.7, 8.4]	[1.1, 1.2]	[0.91, 1.1]

Note: Table displays summary of various statistics (columns) by outcome variable (rows) which are the number of crimes per district-hour. Estimates are recovered from estimating equations (3) for Fixed Effects (FE) district-year-month-day of week -8 hour window fixed effects and equation (4) for first differences using actual (FD) and predicted (PFD) team membership, on three samples (2010-2012, 2013-2015, 2016-2019). The first row for each outcome is the median estimate across sample groups while the second row contains the minimum and maximum across sample groups. All variance terms are KSS bias corrected. Zero or negative values reflects plug-in estimates being entirely attributable to bias.

Table C.4: First Stage Joint F-Statistics and P-Values

Endogenous Variable	FE Type	Shrunken FEs				Raw FEs			
		Pooled	2010-2012	2013-2015	2016-2019	Pooled	2010-2012	2013-2015	2016-2019
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Team Avg. PO FE									
Street Stops	Predicted	4633.2 (p<0.001)	3611.2 (p<0.001)	3345.5 (p<0.001)	2344.4 (p<0.001)	6118.5 (p<0.001)	2450.6 (p<0.001)	4922.3 (p<0.001)	2377 (p<0.001)
Traffic Stops	Predicted	1049.4 (p<0.001)	2370.4 (p<0.001)	3382.9 (p<0.001)	971.5 (p<0.001)	1410.2 (p<0.001)	3217.3 (p<0.001)	3322.2 (p<0.001)	973.2 (p<0.001)
Arrests - Drug	Predicted	3402.1 (p<0.001)	2035.1 (p<0.001)	1212.3 (p<0.001)	1897 (p<0.001)	4235.2 (p<0.001)	3155.4 (p<0.001)	1874 (p<0.001)	1765.1 (p<0.001)
Arrests - Non-Drug Nonindex	Predicted	4875.6 (p<0.001)	2349.1 (p<0.001)	2104 (p<0.001)	1366 (p<0.001)	4523.9 (p<0.001)	2902.1 (p<0.001)	3252.9 (p<0.001)	1246.6 (p<0.001)
Arrests - Violent Index	Predicted	4422.1 (p<0.001)	1760.3 (p<0.001)	3179 (p<0.001)	1180.6 (p<0.001)	4462.7 (p<0.001)	2076.1 (p<0.001)	3304.4 (p<0.001)	1166.4 (p<0.001)
Arrests - Property Index	Predicted	2969.8 (p<0.001)	737.5 (p<0.001)	2621.8 (p<0.001)	998.1 (p<0.001)	3867.9 (p<0.001)	1053.1 (p<0.001)	3059.5 (p<0.001)	1230.6 (p<0.001)
Uses of Force	Predicted	4102.3 (p<0.001)	1487.2 (p<0.001)	2909.2 (p<0.001)	954.3 (p<0.001)	2886.6 (p<0.001)	1875.3 (p<0.001)	1335 (p<0.001)	767.3 (p<0.001)
Complaints	Predicted	2027.9 (p<0.001)	1142.3 (p<0.001)	1856.7 (p<0.001)	1041.3 (p<0.001)	1762.4 (p<0.001)	725.5 (p<0.001)	1567 (p<0.001)	956.9 (p<0.001)
B: Endogenous Variable Only									
N. Index Arrests	Predicted	14.7 (p<0.001)	4 (p=0.046)	6.6 (p=0.01)	3.3 (p=0.068)	18.6 (p<0.001)	7.5 (p=0.006)	7.3 (p=0.007)	3.7 (p=0.055)
N. Index Arrests	Actual	89.9 (p<0.001)	29.3 (p<0.001)	25.2 (p<0.001)	42.2 (p<0.001)	91.1 (p<0.001)	51.5 (p<0.001)	22.2 (p<0.001)	29.7 (p<0.001)
N. Nonindex Arrests	Predicted	93.6 (p<0.001)	35.6 (p<0.001)	29.4 (p<0.001)	35.9 (p<0.001)	77.8 (p<0.001)	27.1 (p<0.001)	30.5 (p<0.001)	24.6 (p<0.001)
N. Nonindex Arrests	Actual	416.4 (p<0.001)	142.7 (p<0.001)	118.5 (p<0.001)	194 (p<0.001)	346.2 (p<0.001)	127.8 (p<0.001)	112.6 (p<0.001)	167.3 (p<0.001)
N. QOL Arrests	Predicted	202.7 (p<0.001)	81.9 (p<0.001)	59.2 (p<0.001)	80.2 (p<0.001)	177 (p<0.001)	73.1 (p<0.001)	73.8 (p<0.001)	38.1 (p<0.001)
N. QOL Arrests	Actual	705.6 (p<0.001)	238.2 (p<0.001)	219.9 (p<0.001)	332.6 (p<0.001)	523.1 (p<0.001)	200.5 (p<0.001)	199.7 (p<0.001)	244 (p<0.001)
N. Street Stops	Predicted	92.8 (p<0.001)	37.1 (p<0.001)	54.5 (p<0.001)	62.7 (p<0.001)	524.8 (p<0.001)	179.2 (p<0.001)	260.2 (p<0.001)	337.2 (p<0.001)
N. Street Stops	Actual	2161.7 (p<0.001)	1035.5 (p<0.001)	818.1 (p<0.001)	1337.8 (p<0.001)	1845 (p<0.001)	1076.6 (p<0.001)	725 (p<0.001)	1106.3 (p<0.001)
N. Traffic Stops	Predicted	93.9 (p<0.001)	52.9 (p<0.001)	36.8 (p<0.001)	65.5 (p<0.001)	552.2 (p<0.001)	423.9 (p<0.001)	249.1 (p<0.001)	355.1 (p<0.001)
N. Traffic Stops	Actual	2525.2 (p<0.001)	1395.6 (p<0.001)	827.4 (p<0.001)	1590.8 (p<0.001)	2272 (p<0.001)	1325.4 (p<0.001)	745.5 (p<0.001)	1390.7 (p<0.001)
C: + Team Avg. FE Controls									
N. Index Arrests	Predicted	3.5 (p=0.002)	2 (p=0.056)	2.6 (p=0.016)	1.3 (p=0.274)	4 (p<0.001)	2.6 (p=0.017)	3.3 (p=0.003)	1.7 (p=0.115)
N. Index Arrests	Actual	76.2 (p<0.001)	24.8 (p<0.001)	21.5 (p<0.001)	31.1 (p<0.001)	77.7 (p<0.001)	41.6 (p<0.001)	18.6 (p<0.001)	21.4 (p<0.001)
N. Nonindex Arrests	Predicted	32.5 (p<0.001)	11.1 (p<0.001)	11.3 (p<0.001)	14.4 (p<0.001)	23.4 (p<0.001)	8.2 (p<0.001)	10.4 (p<0.001)	7.2 (p<0.001)
N. Nonindex Arrests	Actual	168.3 (p<0.001)	68.3 (p<0.001)	53.4 (p<0.001)	62.2 (p<0.001)	163.7 (p<0.001)	119.1 (p<0.001)	88.2 (p<0.001)	135.2 (p<0.001)
N. QOL Arrests	Predicted	39.4 (p<0.001)	15.3 (p<0.001)	12.6 (p<0.001)	17.6 (p<0.001)	31.7 (p<0.001)	13.9 (p<0.001)	14.4 (p<0.001)	8.3 (p<0.001)
N. QOL Arrests	Actual	598.4 (p<0.001)	187.9 (p<0.001)	206.7 (p<0.001)	285.5 (p<0.001)	441.6 (p<0.001)	151.7 (p<0.001)	190.3 (p<0.001)	186.6 (p<0.001)
N. Street Stops	Predicted	92.8 (p<0.001)	37.1 (p<0.001)	54.5 (p<0.001)	62.7 (p<0.001)	79.4 (p<0.001)	32.3 (p<0.001)	53.5 (p<0.001)	43.5 (p<0.001)
N. Street Stops	Actual	2090.8 (p<0.001)	943.7 (p<0.001)	823.3 (p<0.001)	1124.4 (p<0.001)	1689.9 (p<0.001)	953 (p<0.001)	657.2 (p<0.001)	934 (p<0.001)
N. Traffic Stops	Predicted	93.9 (p<0.001)	52.9 (p<0.001)	36.8 (p<0.001)	65.5 (p<0.001)	78.7 (p<0.001)	54.3 (p<0.001)	36.4 (p<0.001)	54.6 (p<0.001)
N. Traffic Stops	Actual	2598.8 (p<0.001)	1351.6 (p<0.001)	807.1 (p<0.001)	1688.2 (p<0.001)	2304.4 (p<0.001)	1254.8 (p<0.001)	708.1 (p<0.001)	1449.3 (p<0.001)

Note: Table displays joint F-test, among excluded instruments only, and associated p-values in parentheses for the first stage regressions, across endogenous variables and specifications.

Table C.5: Map of Incidence Costs to Crime Types using Miller et al. (2021) Table 5

Miller et al. Crime Type	CPD FBI Code	Miller et al. Crime Cost ($)
(1)	(2)	(3)
Murder	01A	7,809,155
Murder	01B	7,809,155
Rape - Police-reported	02	400,793
Robbery - Police-reported	03	43,768
Assault - Police-reported	04A	38,784
Assault - Police-reported	04B	38,784
Burglary - Police-reported	05	5,369
Larceny/theft - Police-reported	06	4,780
Motor vehicle theft - Police-reported	07	10,783
Intimate partner violence	08A	28,251
Intimate partner violence	08B	28,251
Arson	09	39,438
Fraud	10	2,053
Fraud	11	2,053
Fraud	12	2,053
Fraud	13	2,053
Vandalism	14	1,349
Weapons carrying	15	3,725
Prostitution/pandering	16	444
Prostitution/pandering	17	444
Drug possession/sales	18	444
Gambling	19	1,819
Child maltreatment	20	79,320
Drunkenness	22	1,819
Disorderly conduct	24	1,819
Curfew/loitering violations	26	1,819

Note: Table displays the mapping between crime types in the Chicago Police Department (CPD) crime data (Column (2)) to the total incidence costs from crimes in Table 5 of Miller et al. (2021). Not all crime types are discussed in Miller et al. (2021), so cost amounts have been imputed or filled in to close categories, as shown in the table.

Table C.6: Effect of Policing Activity on Crime Costs

	Index Arrests	Nonindex Arrests	QOL Arrests	Street Stops	Traffic Stops
	(1)	(2)	(3)	(4)	(5)
Total Crime Cost ($1,000)					
Main	-51	-19	-18	-4.2	-3.5
	[-280, 180]	[-50, 12]	[-54, 18]	[-11, 2.8]	[-10, 3.1]
Raw FEs	-50	-17	-22	-4.2	-3.6
	[-290, 180]	[-51, 17]	[-63, 18]	[-11, 2.8]	[-11, 3.5]
Predicted Teams	-240	-18	-29	-2.5	6
	[-920, 440]	[-89, 53]	[-110, 48]	[-14, 9.3]	[-9.5, 22]
Main + FE Controls	12	-32	-13	-2	-2.2
	[-230, 260]	[-75, 10]	[-50, 25]	[-9.1, 5]	[-8.8, 4.4]
Total Low-Murder Crime Cost ($1,000)					
Main	-1.4	-1.6	-2.4	-1.1**	-0.24
	[-44, 41]	[-5.7, 2.5]	[-7.1, 2.3]	[-2.2, -0.071]	[-1.3, 0.83]
Raw FEs	-2	-2.2	-4.7*	-1.2**	-0.14
	[-45, 41]	[-6.6, 2.2]	[-9.9, 0.5]	[-2.3, -0.09]	[-1.3, 0.99]
Predicted Teams	-47	-14**	-13**	-0.51	0.69
	[-180, 90]	[-25, -3.2]	[-25, -2.3]	[-2.2, 1.2]	[-3, 4.4]
Main + FE Controls	7.4	3.5	-2	-1.1**	-0.2
	[-38, 53]	[-2.8, 9.8]	[-6.9, 3]	[-2.1, -0.0032]	[-1.3, 0.88]

Note: Table displays the effects different policing activities (columns) on crime costs in $1,000, with 95% confidence intervals displayed in brackets below, over the pooled sample (2010-2019). Estimates and confidence intervals are computed by estimating equation (7) regressing the cost-weighted summation of crimes which occurred in a specific district-hour on the number of policing activities (e.g., street stops) and instrumenting for the policing activity using average team shrunken officer effects for that activity. Estimates can be interpretted as the effect of 1 additional policing activity on cost-weighted crime. Low murder refers to valuing murders at $400,000. Rows within panels adjust the specification: 'Raw FEs' uses unshrunken (raw) officer effects to compute team averages for the first stage; 'Predicted Teams' uses the predicted rotations predict the officers working in that district-hour to construct the team average effects in the first stage; and 'Main + FE Controls' uses the main specification by adds controls in the second stage for team average officer effects on the other activities (e.g., controlling for team average effects for street stops, traffic stops, nonindex arrest, use of force, and complaints if the activity of interest is index arrests). QOL Arrests refer to 'quality of life' arrests, a subset of nonindex arrests. Officer effects are shrunken and computed separately over 2010-2012, 2013-2015, and 2016-2019 using equation (1). Standard errors used to compute confidence intervals are clustered at the year-month-day of week-district-hour level. *** p<0.01, ** p<0.05, * p<0.1.

D Appendix D - Additional Analyses

D.1 Comparison with Arrest Quality

While the quantity of enforcement does not appear to be strongly related to officers' propensities to use force or receive complaints, it is possible that the higher cost officers are making more impactful arrests in terms of quality. As a proxy for higher quality arrests, I classify arrests which resulted in a guilty finding (through trial or a plea) and resulted in the defendant/arrestee being sentenced to incarceration, using data on the Cook County court from the Invisible Institute. However, $\theta^{Incar.}$ measures the officer's effect on the number of high quality arrests, I measure quality as the residual variation in $Var(\theta^{Incar})$ which cannot be explained by arrest quantity.

Following Rose et al. (2022), I use the variance-covariance structure to mimic the regression of high quality arrests on arrest quantity and other activity metrics, as shown in Table D.1. As shown in Columns (1) and (4), for 2010-2012 and 2013-2015 samples, Arrest quantity (using four crime types) explains 67% and 81% of the variation in $Var(\theta^{Incar})$, suggesting that variation in higher quality enforcement is still sizable beyond quantity of enforcement as the standard deviation of the residual variation in θ^{Incar} is around 61% and 54% of the mean of high quality arrests. Columns (2) and (5) add street stop and traffic stop propensities, and Columns (3) and (6) add complaint and use of force. The implied R^2 increases by 0.008 and 0.0005 in Columns (2) and (5), and adding complaints and use of force increases the R^2 by less than 0.003 in both samples. This suggests that beyond the quantity of officer activity, higher cost officers do not have a significant relationship with higher quality of enforcement.

Table D.1: Implied Regression Results of High Quality Arrests on 1 SD in Other Behaviors

	2010-2012			2013-2015		
	(1)	(2)	(3)	(4)	(5)	(6)
Arrests - Drug	0.004	0.004	0.004	0.01	0.01	0.01
Arrests - Nonindex	0.01	0.01	0.01	0.007	0.008	0.008
Arrests - Property Index	0.002	0.002	0.002	0.0008	0.0008	0.001
Arrests - Violent Index	0.002	0.002	0.002	0.005	0.005	0.004
Street Stops	-	-0.002	-0.002	-	-0.0006	-0.0004
Traffic Stops	-	0.0002	0.0001	-	0.0001	0.00002
Complaints	-	-	0.001	-	-	0.001
Uses of Force	-	-	0.0002	-	-	-0.0006
Mean Dep. Var	0.02	0.02	0.02	0.02	0.02	0.02
Change in R2	0.67	0.0082	0.0027	0.81	0.00052	0.0021
SD(Resid PO)/Mean(Y)	0.61	0.6	0.6	0.54	0.54	0.54

Note: Table displays the implied regression results based on the variance-covariance relationship between officer propensities to make arrests which result in incarceration ('high quality' arrests) and their primary behavioral propensities, across the 2010-2012 and 2013-2015 samples. All coefficients are scaled by 1 SD in the respective officer effect. Resid. PO refers to the residualized outcome (officer propensity to make arrests resulting in incarceration) after removing variation explained by the implied regression, and Mean(Y) refers to the mean of the outcome (average number of arrests resulting in incarceration per shift).

D.2 Match Effects

To determine the quality of the fit using the main additive specification in equation (1), I estimate a model using officer-assignment match effects as a comparison. Table D.2 shows that while around 10 - 12,000 parameters were estimated in the main specification, in the match effects model between 63,000 and 80,000 parameters were recovered. Figure D.1 compares the shares of $Var(\tilde{y}^k)$ that are explained by the officer and assignment parameters in the main additive model ($Exp^k = Var(\tilde{y}^k) - Var(\epsilon^k)$,x-axis) compared with the match effects model, either in terms of absolute increase in variation explained ($Exp^k_{Match} - Exp^k_{Add}$, top panel) and percent increase in variation explained relative to the additive model ($\frac{Exp^k_{Match} - Exp^k_{Add}}{Exp^k_{Add}}$) (y-axis).

As shown in Panel A of Figure D.1, estimating over 40,000 additional parameters results

in an increase in share of variation explained of between 0 and 0.09, though this varies by outcome with the rarer and lower variation explained outcomes having smaller increases (complaints, force, index arrests) between 0 and 0.015, nonindex arrests increasing by around 0.02, and stops and traffic stops increasing by around 0.04 and 0.09 after KSS bias correction (left figure). Note that bias correction proves more important for the low frequency outcomes as plug in variation explained (right figure) increases by around 0.02 for force, complaints and index arrests. However, these absolute changes are not proportional to the initial amount of variation explained in the additive model, which is shown in Panel B. While plug in variation explained increases massively in proportion to initial variation explained for rare outcomes (right panel), KSS bias correction shows that in most samples use of force, and index arrests have their variation explained by around 70%, and nonindex arrests, street stops, and traffic stops have theirs increased by around 15%, 30%, and 40%, though complaints have increased theirs by between 60% and 160%, depending on the sample.

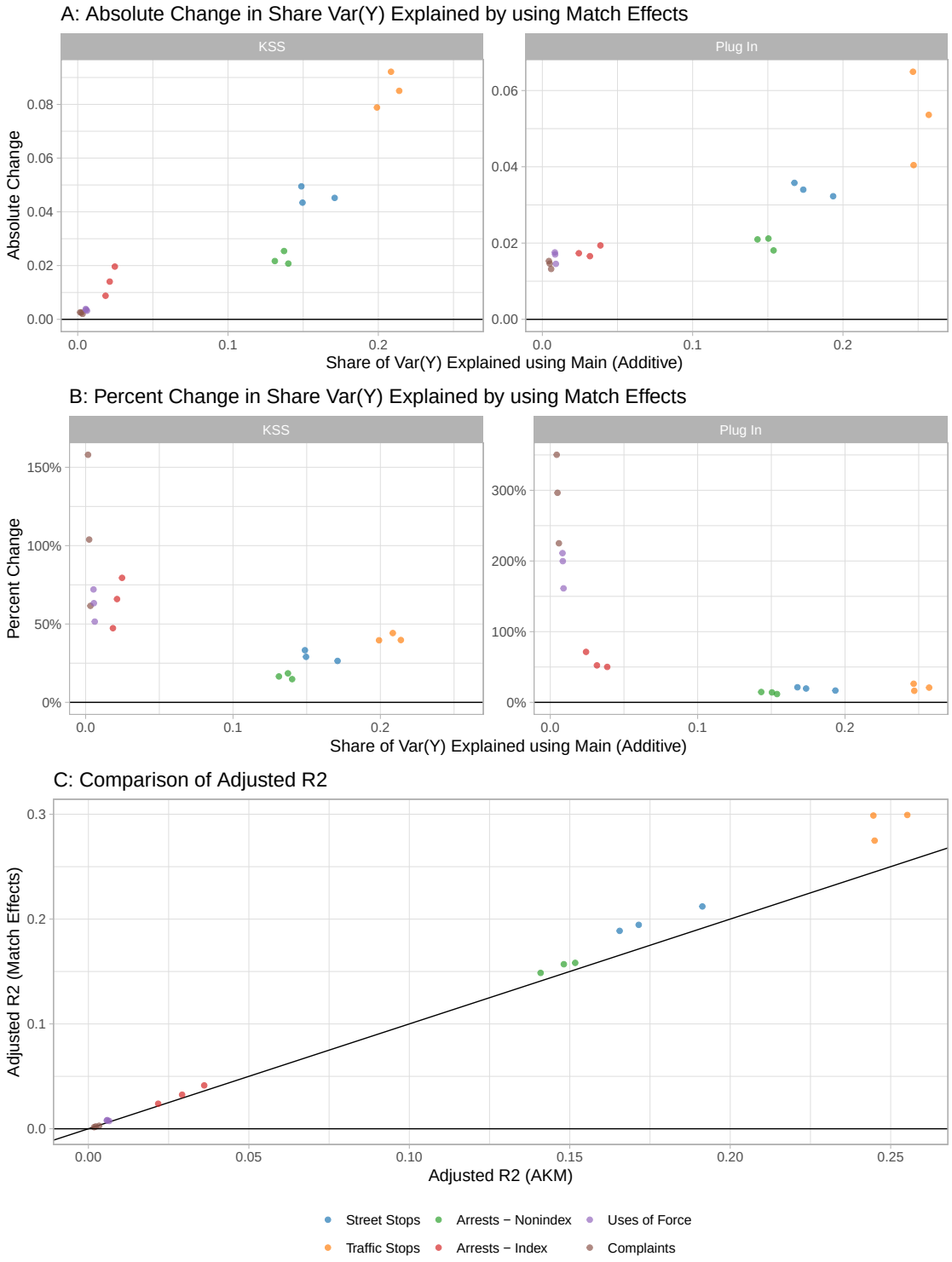
Panel C shows, after taking into account the number of parameters using adjusted R^2, the match effects model does not significantly increase adjusted R^2 for most outcomes. While these results do indicate that match effects between officers and their assignments are empirically important and worth examining, it is clear that the additive model is a useful and parsimonious approximation.

Table D.2: Comparison of Number of Parameters Estimated in Main (Additive) and Match Effects Models

Sample Period	Officer Effects	Assignment Effects	Match Effects
	(1)	(2)	(3)
2010 - 2012	8,618	1,297	63,346
2013 - 2015	8,468	1,437	64,854
2016 - 2019	9,832	1,518	79,328

Note: Table displays the number of parameters of interest (officer effects and assignment effects) estimated in each sample period in the main specification (Columns (1)-(2)) while Column (3) show the number of officer-assignment match effects estimated when interacting officer and assignment effects.

Figure D.1: Comparison of Variation Explained in Additive (AKM) and Match Effects Model



Note: Figure summarizes comparisons in the amount of variation explained using the additive AKM model (equation (1)) compared with using a match effects model (officer-assignment effects). Panel A displays the share of variation in the outcome explained by the AKM model (x-axis) compared with the absolute change in share of variation explained when using the match effects model (match effects − AKM); Panel B has the same x-axis, but the y-axis is the percent change in share of variation explained (i.e., y-axis in Panel A divided by the x-axis). In Panels A and B, the first figure ("KSS") shows the KSS bias corrected results and the second figure ("Plug In") shows the non-bias corrected results. Panel C compares the adjusted R^2 for the AKM model (x-axis) to that of the match effects model (y-axis) and the solid black line denotes the 45-degree line. Each point represents the result from a specific outcome (color) and sample period.

D.3 Overtime

Officers earn overtime (generally 1.5 times their hourly rate) for any work after 40 hours a week, which can occur if the officer appears in court, works an extra event, or is completing an arrest which spills over beyond their regular shift. Naturally, this can create an incentive for officers to sort into assignments where they can make more overtime, and it is possible that overtime effectively extends the officers' shift which is not included in the primary analysis.

I employ additional data on overtime shifts from 2014 to 2019 (used in the 2016-2019 sample period only due to completeness) to study the relationship between overtime and the prior results. I create a new outcome being the number of hours an officer works as overtime of various types. At the daily level, an officer's overtime hours are counted for their shift if they worked the overtime immediately following their shift (or overlapping with it). Because officers may also just work overtime on the next day, for things like court appearances following an arrest, I also aggregate the data at the weekly level and I count overtime as any overtime worked in that week, regardless of whether it was worked immediately after a regular shift.

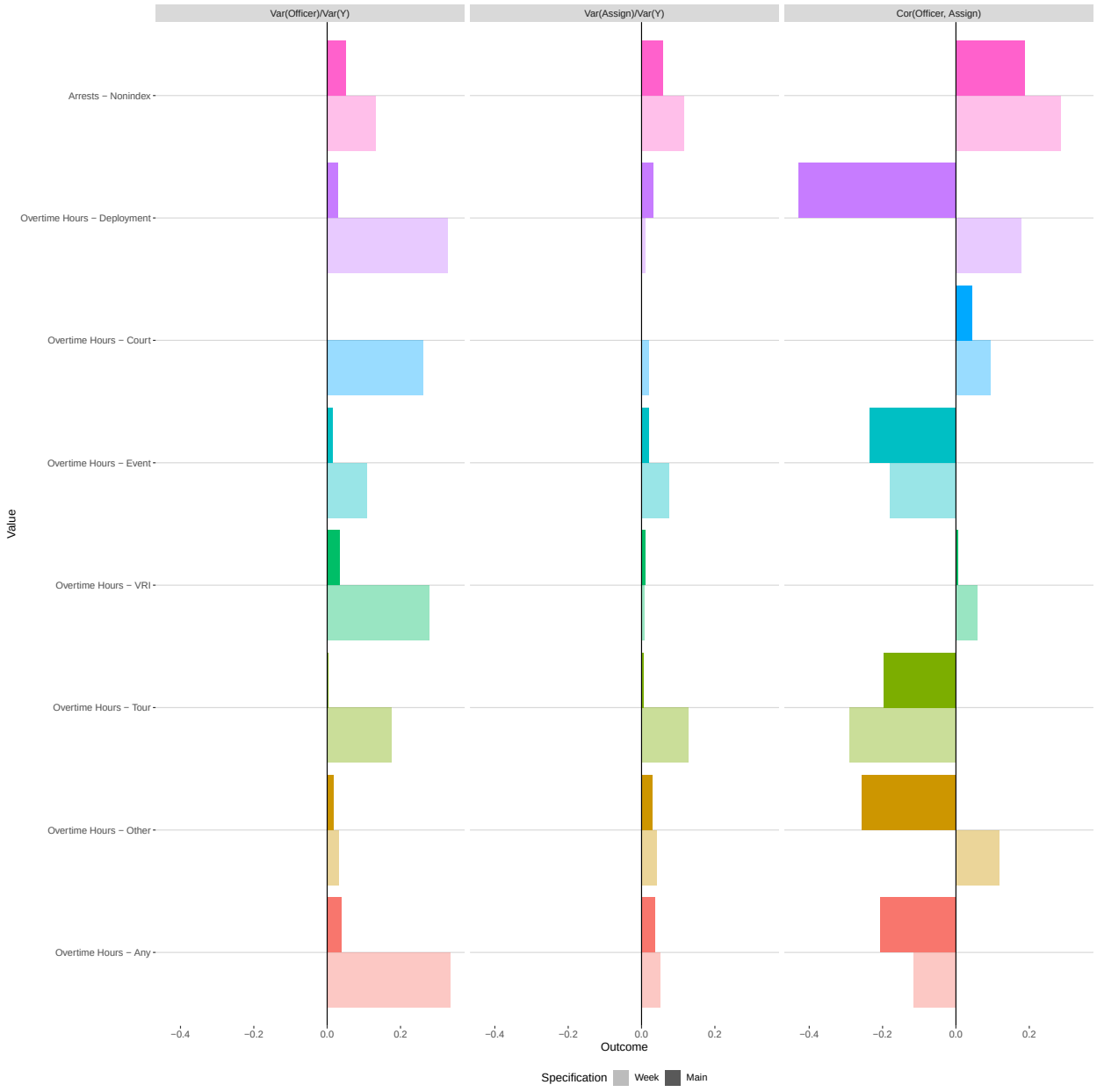
Figure D.2 displays the share of variance in the number of overtime hours worked that is explained by officer and assignment effects and the correlation between officer and assignment effects for each overtime type and non-index arrests (for comparison). At the daily shift level, officer and assignment effects explain virtually none of the variation in overtime for court, but explain small amounts of variation for other types of overtime, and a small portions of overtime for deployments and VRI (violent reduction patrols). Additionally, at the shift-level, there are strong negative correlations between officer and assignment effects across overtime types. However, at the week-level, variation in officer effects explain sizable portions of the variation in overtime for deployments, court, events, VRI, and additional tours, while assignment effects explain relatively small portions. Additionally, while correlations between assignment and officer effects are generally negative (with the exception of court overtime which is small and positive), the negative correlations are much smaller than at the daily

level.

These patterns indicate that officers do vary significantly in their desire to work overtime for additional policing and appearing in court, but this is generally not driven by the extension of regular shifts, but rather working additional overtime on different days or times disconnected from their regular shift. Additionally, officers are not clearly sorting into assignments which permit additional overtime, with the exception of overtime from court (at the weekly-level) consistent with officers sorting into assignments that allow for more arrests then those arrests leading to more court overtime.

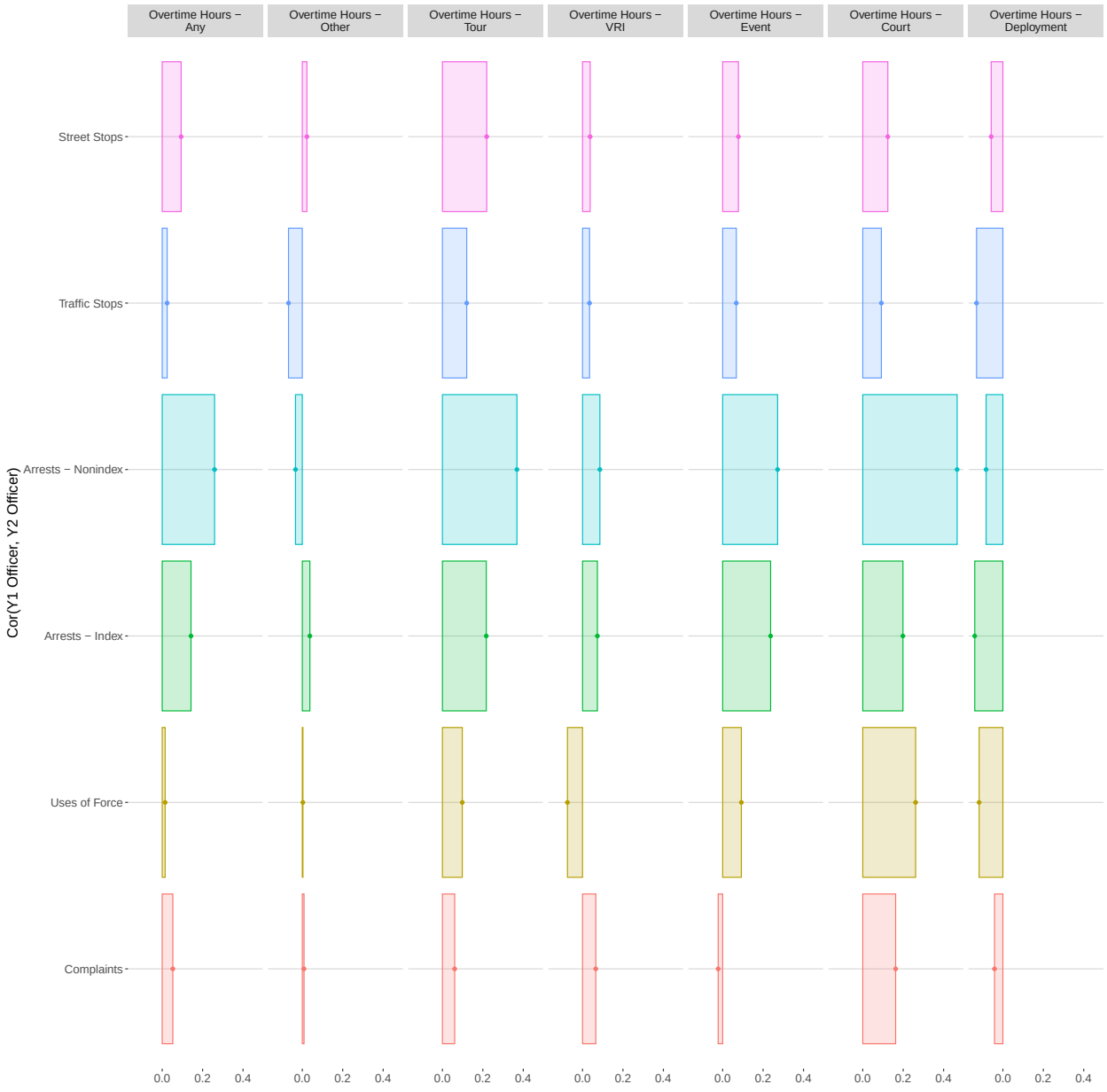
Next, given that overtime is a product of officers working additional hours, we can determine if the officers who work more overtime are productive during their regular shifts. Figure D.3 displays the cross-outcome correlations in officer effects with the other main outcomes using the weekly-level data. Overtime for activities such as deployments and VRI have only weak relationships with the primary activities. However, overtime for court, events, and tours are correlated with all primary activities. Court and tour overtime is most strongly correlated with nonindex arrests though only at around 0.4, while event overtime is less strongly correlated with such metrics though index and nonindex arrests are the largest. Overall, officers who are more active do work more overtime, and some of this is driven by arrests leading to court overtime, however variation in officers' arrest propensity by no means explains all of the variation in their overtime propensities.

Figure D.2: Decomposition of Variation in Officer Overtime



Note: Figure displays the KSS bias corrected variance and correlation component from estimating equation (1) on officer overtime in the 2016-2019 sample, with the same statistics for nonindex arrests displayed for comparison. Shading denotes the specification, either the main specification using shift-level data or data aggregated to the week level, though both use 2016-2019 sample.

Figure D.3: Correlation between Officer Effects on Main Outcomes and Overtime Hours



Note: Figure displays the median KSS bias corrected correlation between officer effects on officer-level outcomes (rows) and overtime hours worked by the officer (columns) for the 2016-2019 sample period from estimating equation (1).

D.4 Alternative Team Specifications

We can probe the robustness of these results and quality of the additive specification by employing alternative empirical strategies. First, there is the concern that using a single

effect for each officer (ω_i) causes significant misspecification because the officers may work in different roles. Similarly, it is possible that officers may be correlated with their supervisors, i.e., the sergeants on duty, which are actually driving the effects.

To account for these potential issues, I extend the model to allow for two teams — e.g., one for regular officers and one for tactical team officers, one for officers and one for sergeants — such that:

$$\Omega^1_{\tau(d,h)} + \Omega^2_{\tau(d,h)} = \frac{1}{N^1_{\tau(d,h)}} \sum_{i \in \tau^1(d,h)}^{N^1_{\tau^1(d,h)}} \omega^k_{i,1} + \frac{1}{N^2_{\tau(d,h)}} \sum_{i \in \tau^2(d,h)}^{N^2_{\tau^2(d,h)}} \omega^k_{i,2} \tag{9}$$

So, in the regular vs. tactical team case, each officer in the original specification can have two ω_i's, one for when they work as a beat officer and one for when they work as a 'special' (non-beat) officer. In the case of adding sergeant controls, this simply involves constructing a separate 'team' of the sergeants on duty.

If the initial model is misspecified, then we expect that after bias correction, the total amount of variation explained in the model and the variance attributable to the officer components (after bias correcting all variance and covariance components) should increase. For example, we expect that given

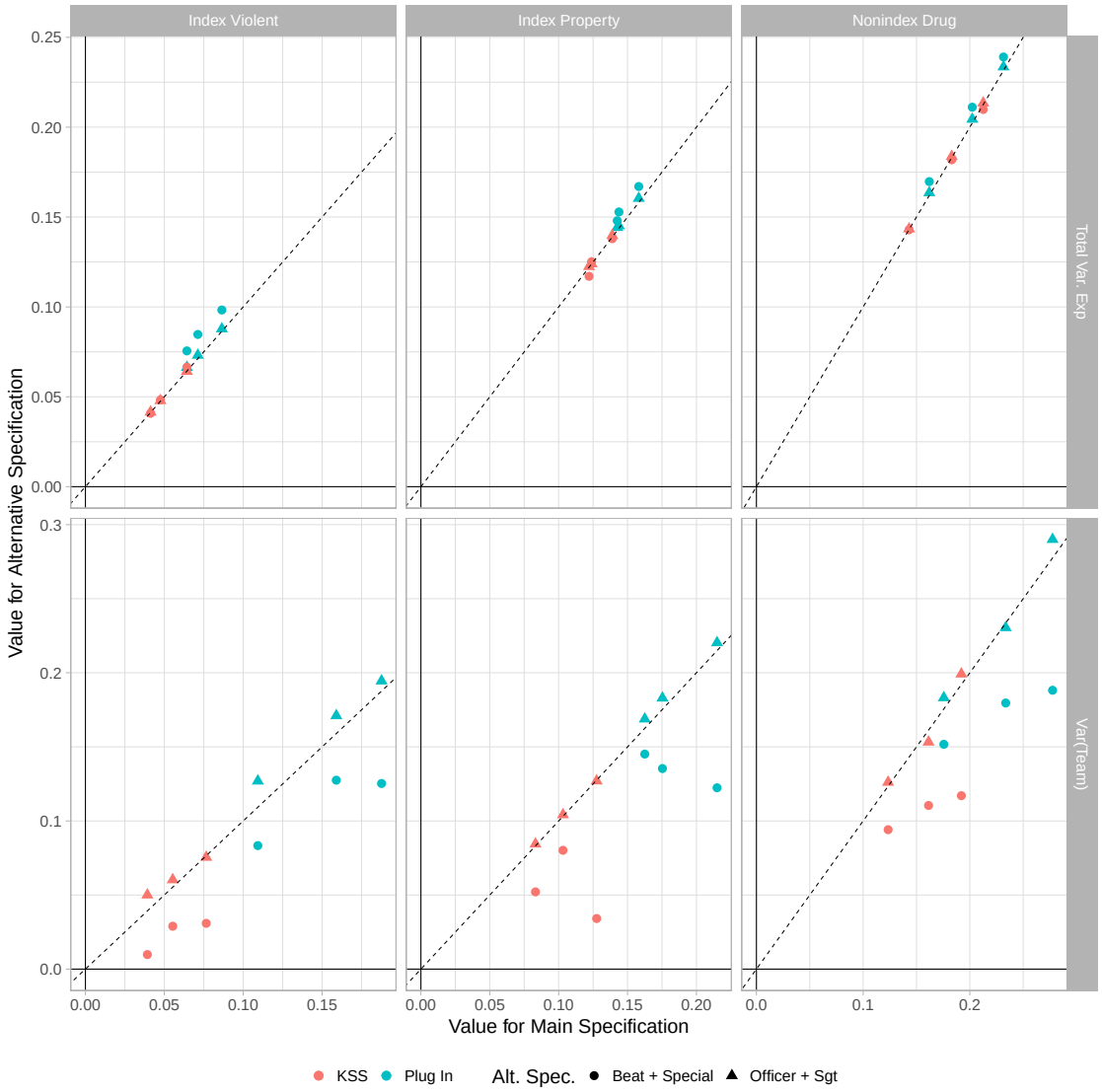
$$Var(Y) = \underbrace{Var(\Omega^1) + Var(\Omega^2) + Var(\gamma) + 2Cov(\Omega^1, \Omega^2)}_{Var(\Omega)} + 2[Cov(\Omega^1, \gamma) + Cov(\Omega^2, \gamma)] + Var(\epsilon)$$

then $Var(\Omega)$ should be larger than $Var(\Omega)$ recovered from the model in which there is no additional officer type and total variation explained by $(Var(Y) - Var(\epsilon))$ should also be larger with more officer types.

Figure D.4 displays both total variation explained $(Var(Y) - Var(\epsilon))$ and the total variation of officer components $(Var(\Omega))$ compared with these values in the main specification with a single officer type. As expected given the larger number of parameters, the multi-officer model has higher variation explained before bias correction. After bias correction, there is little evidence that the multi-officer models explain more variation than the single

officer model. Additionally, the total variation in the officer components is not consistently and significantly larger after bias correction, suggesting that while multi-officer models can increase variation explained it the single officer model is more efficient given finite data and not significantly misspecified.

Figure D.4: Comparison between Main and Alternative Team Constructions



Note: Figure displays comparison of total variation explained and the variance of the team composition ($Var(\Omega)$) when using the single officer type fixed effects specification on the x-axis with the same estimates when using alternative team constructions. Colors denote whether the estimate is KSS bias corrected or not ('Plug In'). Shapes denote the alternative team constructions. Each point corresponds to a specification with or without KSS bias correction in a specific sample (multi-year periods). All specifications use district-year-quarter-day of week-8 hour period fixed effects due to increased number of parameters in the alternative specifications (whereas the original main specification uses month not quarter).

D.5 Cross-Crime Correlations

We can recover the correlations between team effects on different crime types, as shown in Tables C.1-C.2. Violent crime effects are correlated more strongly with effects on simple violent crime (0.45), and calls relating to domestic disturbances, assaults, and robberies (0.51, 0.4, 0.43) — similarly team effects on robberies and assaults are most strongly correlated with effects on 911 calls about robberies and assaults. Additionally, violent crime effects are positively but not very strongly correlated with property crime effects (0.26). Property crime effects are most strongly correlated with other nonindex offenses, which contains vandalism and nonindex property crimes. Notably, team effects on theft and MV theft are positively, but not strongly, correlated with corresponding calls for 911 calls (between 0.15 and 0.27). In general, team effects tend to be positively correlated across crime types, suggesting teams that are better at reducing one type of crime are not, instead, increasing another type of crime.

Interestingly, team effects on drug crime, which are driven entirely by discretionary arrests, are relatively weakly but positively correlated with other crime types, suggesting teams which are 'good' at making more drug arrests are not also reducing other types of crime through contemporaneous deterrence or incapacitation. Rather, drug crime effects are most strongly correlated with other 'discretionary' crime, in particular, weapons crimes with a median correlation of 0.39.

D.6 Settlement Costs

I use data on settlements due to actions of Chicago Police officers from 2010 to 2019. This provides a link between complaints and monetary costs, similar to Rozema & Schanzenbach (2019) though their focus is predicting future settlement costs while mine is to look at the contemporaneous association between complaints of misconduct and settlement costs. The settlement data was obtained from the Invisible Institute.

Between 2010 and 2019, there are 978 cases with settlement amounts ranging from less

than $6,000 almost $20 million, involving around 2,500 officers — totalling 3,594 officer-settlement observations. I group the settlements by sample period (2010-2012, 2013-2015, 2016-2019) and divide the total settlement amount by the number of officers involved (averaging 7.7 officers), then I compute officer-by-period totals for settlement costs and settlement costs capped at $100,000 per officer to reduce influence of outliers. I link this with the shrunken fixed effects for complaints and total officer observations during the period, and I impute $0 in settlements for officers who are not in the settlement data and drop officers who are in the settlements but not in the analysis sample (and officers must have at least 100 analysis sample observations). The result is 24,274 officer-period observations, with 9.1% of observations having non-zero settlements but averaging $5,347 in settlement costs and $1,850 in capped settlement costs.

Table D.3: Association between Officer Complaint Effects and Settlement Costs

	Settlement Cost per Shift		Settlement Cost	
	(1)	(2)	(3)	(4)
Officer Complaint FE	12190.6***	3572.5***	5379333.8***	1701157.5***
	(542.3)	(241)	(234860.1)	(166276.1)
R2	0.002	0.046	0.0037	0.064
N. Obs	24274	24274	24274	24274
Capped Costs	-	X	-	X

Note: Table displays results of regressing the dollar amount of settlements per named officer in which an officer was named on their shrunken fixed effect for complaints within a sample period. Columns (1) and (2) use total settlement cost (per named officer) by the number of shifts worked for the dependent variable and Columns (3) and (4) use total settlement cost (per named officer) over the sample period. Officers with no settlements and with estimated fixed effects are included with 0s for their settlement costs. Capped costs refers to capping settlement costs per officer at $100,000. All specifications include sample period fixed effects and a linear control for total shifts worked. Standard errors are clustered at the sample period level.

To obtain estimates for the dollar cost associated with an officer causing 1 additional complaint per shift, I divide the officer's settlement costs by their shifts worked, then regress

costs per shift on shrunken officer effects for complaints and total observation controls and sample period fixed effects. Table D.3 displays the results. Using non-capped costs, officers who make 1 additional complaint per shift incur settlement costs $12,190 per shift and $3,572 per shift using capped costs. This is consistent with regressing total settlement amounts (not per shift) on officer complaint effects, with produces estimates of $5.3 million (uncapped) and $1.7 million (capped) per 1 unit increase in complaint effects per sample period (generally around 500 shifts).

E Appendix E - Data

E.1 Assignment and Attendance Data

The primary data for this study is daily assignment and attendance records for all units in the CPD between 2009 and 2019. Data for 2009 was largely unused since multiple outcomes only start in 2010 and institutional features shifted. This data contains officer-by-shift information, where a shift is a working spell by the officer generally for around 9 hours. This data indicates whether the officer was or was not present for duty, the reason for their absence (e.g., it was their regular day off, a vacation day, they were in training), their start and end times, their title (e.g., police officer, sergeant, detective), and their assigned beat, watch, and unit, among other information.

Initially, this data contains over 35 million entries including days-off. I restrict my sample to days where the officer is present for duty with a known assignment and to the sample to officers who are the lowest and most common rank of "police officer" (around 80% of the CPD). They must also be on regular assignments (e.g., not on desk duty, training, SWAT teams, or serving on lockup). I remove all observations from units in which are uncommon and where officers do not do much measurable police work (e.g., no arrests occur in training units), and I remove all officers who could not be matched to other data sets. Finally, I remove probationary officers or those in the academy (less than 1.5 years of tenure) and a small number of observations with officers with reported tenure of 30 years or more. These filters result in 13,928,210 officer-shift observations, with 13,400 unique officers, in 54 units (25 of which are patrol units that correspond to districts).

E.2 Officer Behavior and Performance

To measure the many dimensions of officer-level behavior and performance, I collected data many of the officer-level activities considered in policy and academic research: arrests, stops, use of force, and misconduct. I link the daily assignment records to these officer activities.

Panel C of Table 1 summarizes the means of the primary outcomes per 100 shifts, while Figure E.1 displays the average outcomes per 100 shifts by year. Table E.1 displays means and standard deviations for the primary and supplementary outcomes per 100 shifts.

E.2.1 Street and Traffic Stops

First, the most common outcomes are 'street stops' and 'traffic stops'. Both types of stops can be considered a metric for contact with suspects and non-suspects alike, and they are often used in "broken windows" and proactive policing to measure officer performance, as was the case in Chicago under Superintendent McCarthy (2011 - 2015). In 2011, Garry McCarthy started as the Superintendent of Police in Chicago. This is notable because McCarthy brought a different perspective on policing from his experience under Bill Bratton in the New York Police Department, which involved implementing CompStat and following "broken windows" policing (Moser (2012)). This resulted in an emphasis on proactive policing and measuring performance, and the CPD began relying heavily on stop-and-frisk (Skogan (2022)). In 2012, under McCarthy, there was a districts reorganization with three smaller districts being dissolved and merging with larger districts.

The data for both street and traffic stops starts in 2010 in the records obtained for this project. 'Street stops' refers to contact cards (CC), which were used from 2010-2015, and investigator stop reports (ISR), which were used from 2016-2019 excluding those classified as 'traffic'. In the sample of shifts, there are 35 street stops per 100 shifts. In late 2015, following a report on CPD stop and frisk (ACLU (2015)) by the ACLU, the CPD agreed to collect additional information on their stops replacing CC with ISR, which took effect on January 1st, 2016. As in Figure E.1, street stops rose significantly from 2010 to 2015 from just around 30 to over 60 per 100 shifts. However, in 2016, stops declined massively to less than 15 per 100 shifts and increased only modestly to just less than 20 per 100 shifts in 2019.

'Traffic stops' are the second most common outcome, with an overall mean of 12.5 per 100 shifts. This data was obtained separately FOIA-ing for traffic stops rather than contact

cards, and the CPD often categorizes them separately. Traffic stops became significantly more common in 2016, rising from less than 10 per 100 shifts between 2010 and 2015, to 20 in 2017 and around 35 by 2019. As discussed in Hausman & Kronick (2023), this shift occurred because the ISR reform removed 'pedestrian' stops (CC then ISR) as a feasible activity (ISR cards were significantly more involved) to be used as a performance metric or target suspects. (Note that while a 'stop' can be attributed to up to two officers, the traffic stop data only includes on officer per event.)

E.2.2 Arrests

Next, I link data on officers' arrests from 2009 to 2019. Arrests are the primary metric for enforcement and are often used as a proxy for productivity. Arrest data include both arrestee information, crime types, and arresting officer type (primary, secondary, and auxiliary/supporting). In the main results, I credit an officer for an arrest if they are listed as a primary or secondary arresting officer, and I divide arrests into two general categories: arrests for index crimes, 'index arrests', (violent: murder, robbery, criminal sexual assault, aggravated assault; property: burglary, motor vehicle theft, larceny/theft, arson) and arrests for non-index crimes, 'nonindex arrests', (e.g., drugs, prostitution, simple battery, vandalism, municipal code violations, traffic offenses). Nonindex arrests are considered more discretionary and often do not have a victim, whereas all index crimes have a victim and are considered less discretionary.

Index arrests are less common on average than nonindex arrests, at around 1.6 per 100 shifts compared with 11 per 100 shifts, and follow different patterns over time. Index arrests decline from around 2 to 1.5 per 100 shifts between 2009 and 2011, then are relatively stable until 2017 after which they decline slightly. In contrast, nonindex arrests steadily decline from 15 per 100 shifts in 2009 to just under 10 per 100 shifts in 2015; in 2016 they significantly drop to around 7 per 100 shifts and trend upward from there until 2019.

Figure E.2 displays the trends in index and nonindex arrests and their subtypes over time.

Total arrests have fallen over the decade in the sample. Index arrests have fallen by around 25%. However, this is due to high numbers during 2009, but from 2011 to 2019 they have not changed significantly, declining by around 15% for property arrests and no real change for violent arrests. Nonindex arrests have fallen consistently over the time period, by around 30%. While arrests for warrant and simple violent crimes have stayed relatively constant, the overall decline has been driven by declines in drug, municipal, and other nonindex offenses, while weapon and traffic arrests have risen slightly.

E.2.3 Use of Force and Misconduct

Beyond contacts and enforcement, police use of force and misconduct are notable metrics for negative outcomes which policy attempt to mitigate and avoid whenever possible. Uses of force by officers are collected using Tactical Response Reports (TRR) data from 2009 to 2019 which contain all instances when officers report using force on a subject, ranging from relatively minor force (e.g., pressure points, defensive techniques) to gunshots and tasers (which are relatively rare). Uses of force occur at around 0.36 per 100 shifts, but were close to 0.4 between 2009 and 2012 the declined until 2018 to around 0.25 with the sharpest drops being after 2012, 2015, and 2017. See Holz et al. (2023) for more details on use of force in the CPD.

Officer misconduct is measured using complaint data which contains complaints filed against police officers by either civilians or fellow officers — though civilian complaints comprise the vast majority of complaints ($> 85\%$). The data on complaints is sourced from two different oversight regimes. The complaint data which covers 2009 to 2015 (in full) was obtained from Independent Police Review Authority (IPRA) in November of 2016 by the Invisible Institute, occurring at a mean of 0.24 per 100 shifts between 2009 and 2015.

In 2017 IPRA was largely replaced by Civilian Office of Police Accountability (COPA) following the post-Laquan McDonald reforms. In late 2015, the Laquan McDonald scandal began following the release of the video of his murder at the hands of CPD officer Jason

Van Dyke. This resulted in significant political changes in Chicago, a DOJ investigation and report, and the major reform of the oversight agency in Chicago among other changes. The change in oversight agencies resulted in a shift in the complaints data as the different regimes produced different data — see Rivera & Ba (2023) for more information. I obtained an extended data set on complaints from the CPD which contains the full breadth of the data. The 2019 data include fewer complaints overall than the 2016 data, with a mean of 0.096 per 100 shifts. The dynamics differ across the data sets, but both decline from 2009 to 2016, and the 2019 complaints increase slightly around 0.06 to 0.09 per 100 shifts after.

E.3 Police Teams

To connect officers working at a given time in a given place with the outcomes that are not directly observed at the officer-level (e.g., crimes), I construct teams of police officers who are working together within the same location (district) at the same time (hour). Police teams are constructed using the raw assignment data with similar filters but with important differences. The unit of observation is the district-hour (e.g., district 12 on June 15th, 2013, at 2 pm), and I allocate officers to 'teams' (the set of officers working in a specific district at a specific hour) based on the start and end times of their shifts when they are present for duty, which is recorded in the assignment data.

I restrict the analysis to only the geographic units, which map directly to police districts (1-25), and count officers as active if they are present for duty at a specific time based on their start and end times (rounded to the nearest hour). I remove officers who do not have reported start or end times, and include observations officers working overtime (only available for 2014-2019) but within the unit. I remove officers who are reported to be in court, meetings, lockup, absent, desk officers, or working as an field training officer, but I record total manpower to use as a control (e.g., number of desk officers). To ensure sufficient observations, I remove a small number of officers within a year that have fewer than 40 observations (hours), and I remove a small number of unit-months which do not have sufficient

observations, days which have too few observations, and all observations in the months where units were collapsed. For team composition, I only count officers working as police officers (the focus of the prior sample), though I do record the number of other ranks active by unit-hour to use as controls (e.g., number of detectives).

Most officers are regular officers rather than tactical team officers (40 vs. around 10). 4 officers are assigned to desks on average, and while others are working overtime of some sort. On average there are very few detectives working in these patrol districts, though there are on average 6 sergeants and 3-4 other non-police officers working in a given hour, and 4 officers excluded from the team due to tenure (e.g., probationary officers).

E.4 Crime

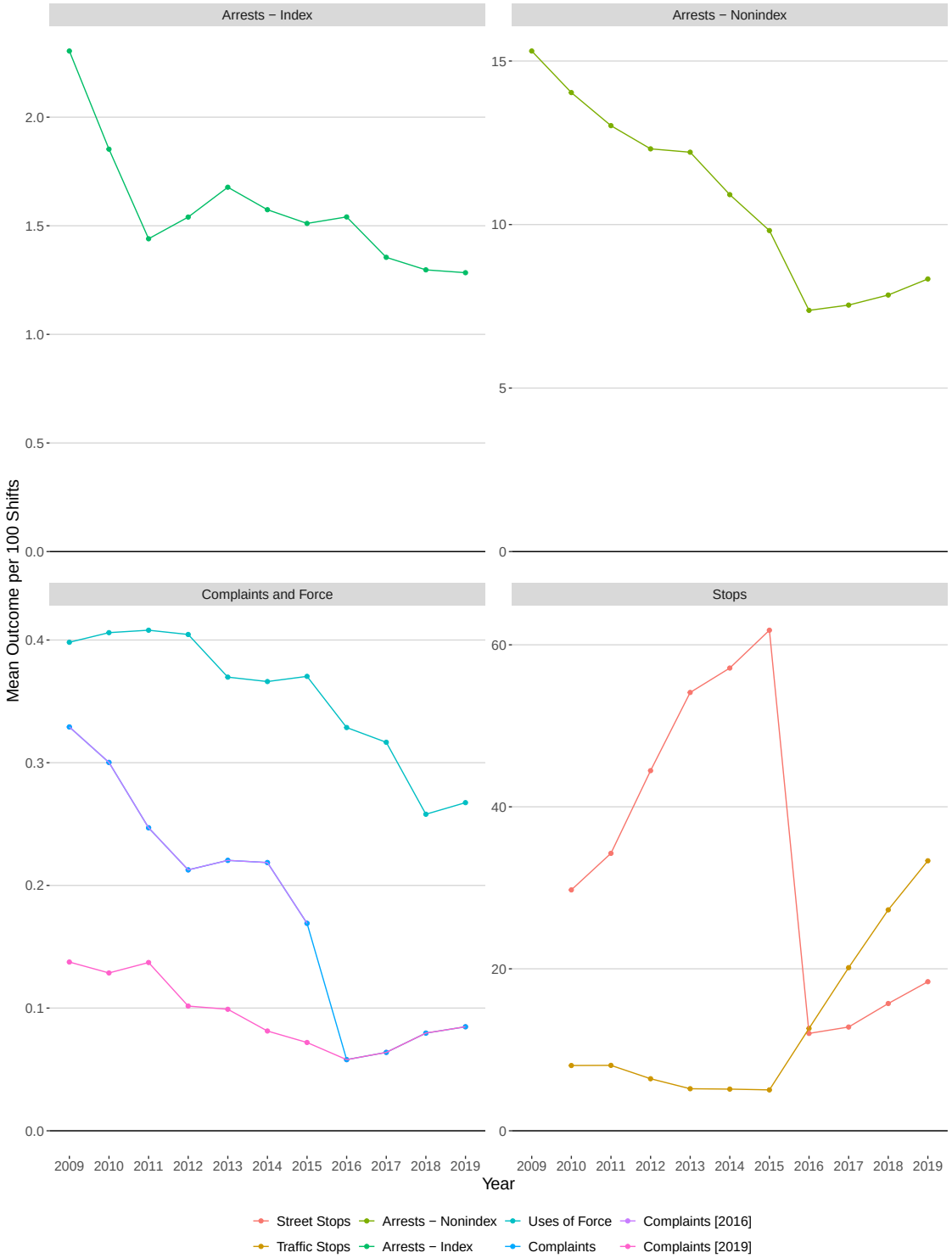
Reported crime is primary outcome of interest which cannot be directly assigned to an individual officer but can be linked to a specific place and time. Geo-located and time-stamped crime data were obtained from Chicago's public data portal. This data contains information on reported crimes in Chicago including the district, coordinates, date and time, whether an arrest was made, crime type, and other information.[22]

Table A.1 displays the average crimes per district-hour broken into cleared (an arrest was made) and uncleared (no arrest was made) crimes with corresponding clearance rates. Drug crimes, with around 0.15 occurring every hour in the average district, are almost all cleared, consistent with them generally being 'victimless' crimes that civilians have little incentive to report, so reports of drug crime come almost all from arrests made at the discretion of officers. Weapon offenses, which occur at a rate of 0.02 per district-hour, are also generally cleared by arrests at 77%, for similar reasons as with drug crime though they are more likely

[22]Reflecting broader trends in the U.S., the environment around crime and policing in Chicago changed significantly between 2009 and 2019. Figure E.3 displays the trends for reported crimes in Chicago over the time period broken into index and nonindex crimes and corresponding clearance rates. Crime across categories declined significantly between 2009 and 2012, while clearance rates remained relatively stable. From 2013 to 2015, crimes also declined but more modestly, while clearance rates increased by a small amount for index crimes. Following 2015, nonindex crimes continued to decline and plateau and index property crimes saw a modest increase but index violent crimes increased significantly into 2017 and declined after; clearance rates during this crime increase declined across crime types and only increased slightly following 2017.

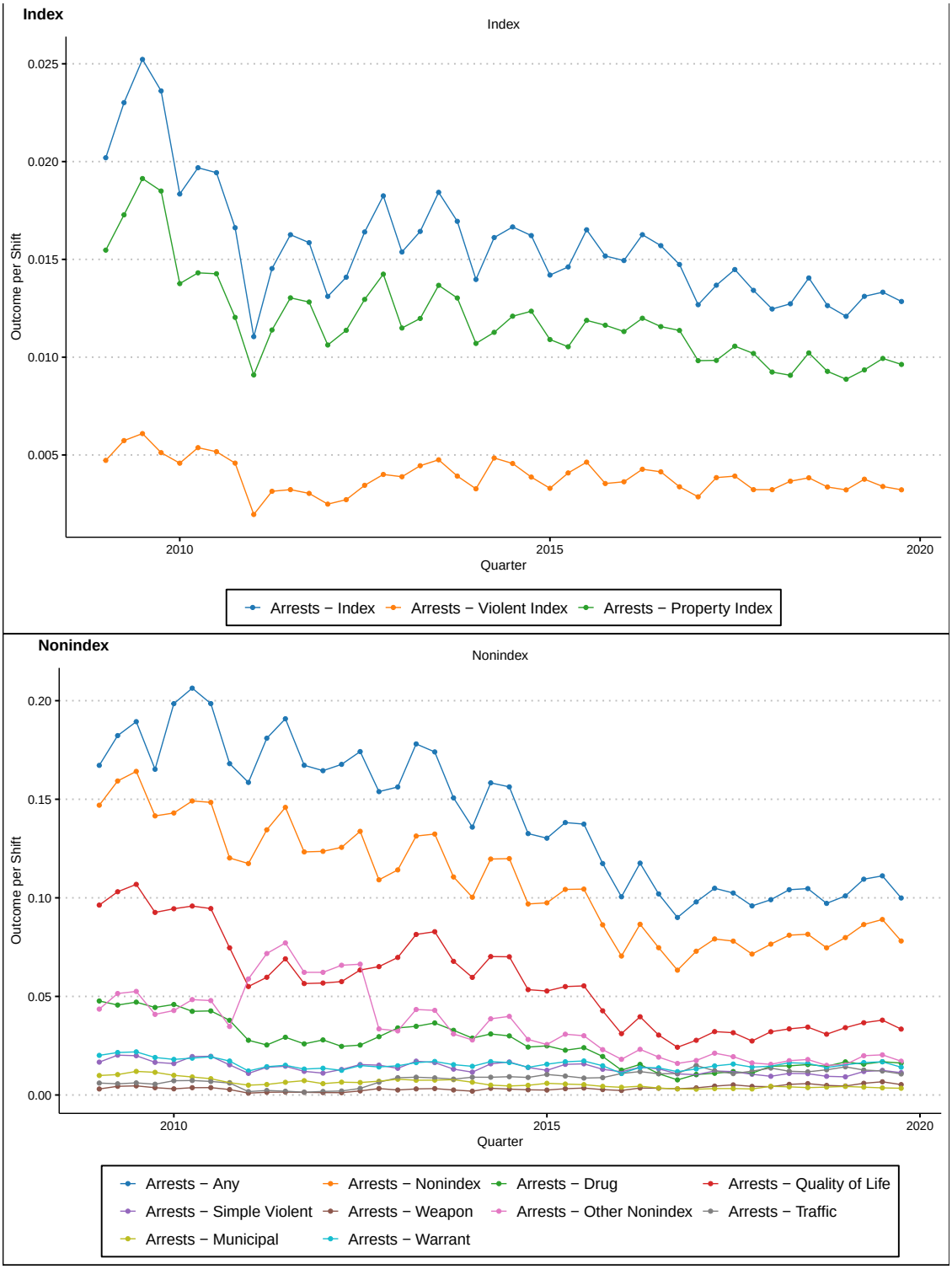
to be reported by civilians than drug crimes. Clearance rates are significantly lower for other crime types. For example, simple violent offenses (nonindex), which occur at around 0.31 per district-hour and consist of simple assault and battery, often due to domestic violence, are only cleared 22% of the time, with a similar rate of other nonindex crimes (occurring at 0.43 per district-hour). Index crimes are less likely to be cleared with arrests. Only 17% of violent index crimes (murder, criminal sexual assault, robbery, rape, and aggravated assault) are cleared with arrests and they occur at a rate of 0.14 per district-hour. Index property crimes (motor vehicle theft, theft, burglary, arson) are rarely cleared at only 9% but are fairly common at almost 0.5 per district-hour.

Figure E.1: Average Officer-Level Outcomes per 100 Shifts over Time



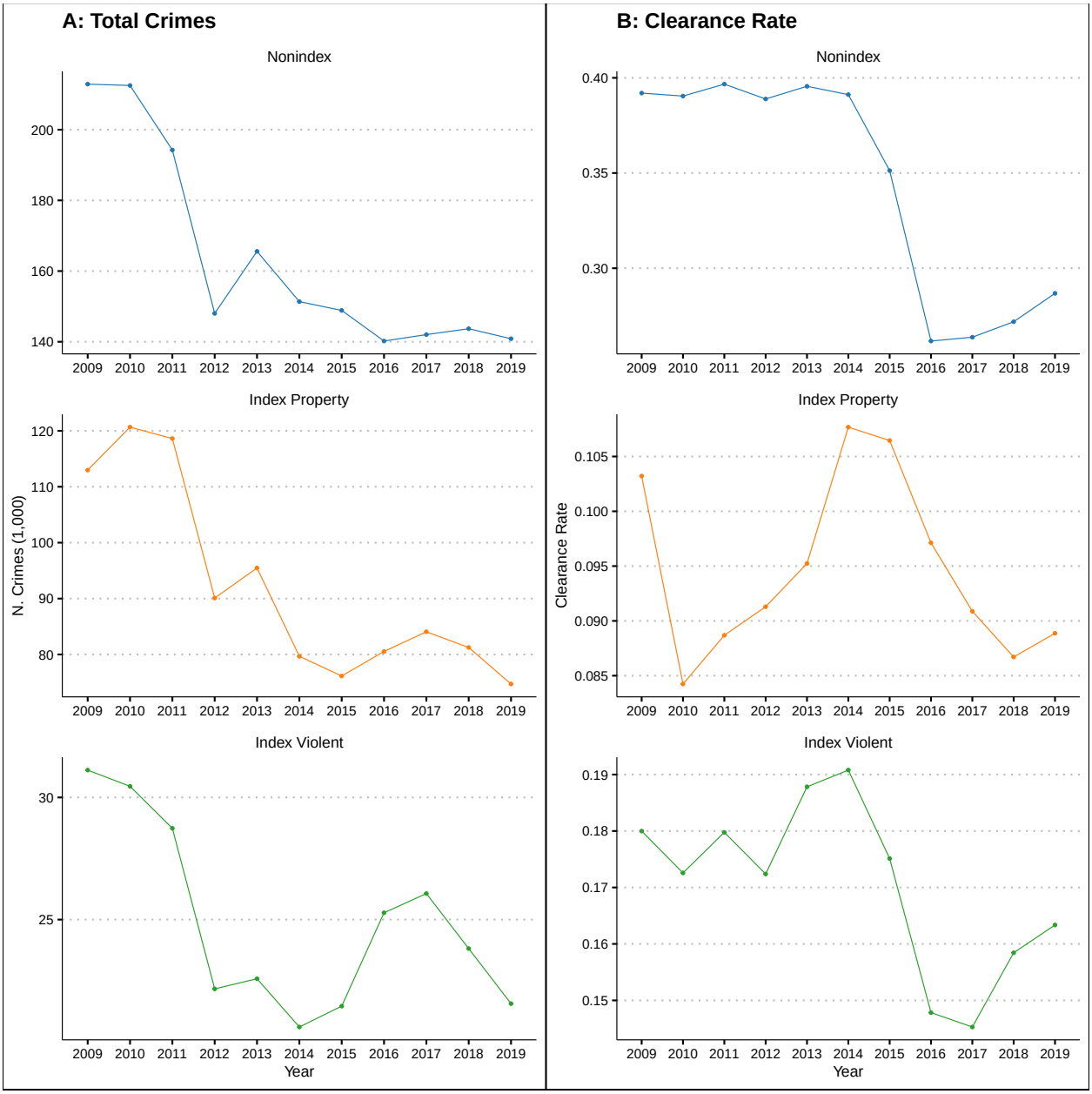
Note: Figure displays the average officer-level outcomes per 100 shifts by year. Street refers to non-traffic contact card (2010 - 2015) and ISR (2016-2019), and complaints [2016] end in 2016 while complaints [2019] end after 2019.

Figure E.2: Timeseries of Arrests by Subcategory



Note: Figure displays the timeseries of mean arrests per 100 shifts of the officer-day data with panels for index arrests and nonindex arrests broken into subtypes at the year-quarter level.

Figure E.3: Timeseries of Crime and Clearances



Note: Figure displays the total number of crimes in 1,000 reported in Chicago between 2009 and 2019 by year broken into subtypes denoted by color (index violent, index property, and nonindex) in Panel A and the clearance rate (share of crimes was reported to have been cleared by an arrest) in Panel B.

Table E.1: Means and Standard Deviations of All Outcome per 100 Shifts by Sample Period

	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)
Stops	51.06 (136.41)	77.21 (175.64)	14.66 (60.85)
Street Stops	35.89 (110.85)	57.23 (147.94)	14.66 (60.85)
Stops - Traffic	15.17 (68.59)	19.98 (83.59)	0 (0)
Traffic Stops	7.42 (41.94)	5.11 (36.84)	23.23 (75.42)
Traffic Stops - Search	0.47 (7.38)	0.25 (5.22)	0.24 (5.12)
Traffic Stops - Ticket	4.34 (32.29)	3.73 (32.35)	3.27 (26.2)
Stop - Traffic (Primary)	11.65 (57.03)	13.27 (64.64)	0 (0)
Stop - Street (Primary)	19.43 (75.7)	30.67 (101.75)	8.07 (41.63)
Arrests - Any	17.59 (53.18)	14.49 (46.28)	10.21 (35.93)
Arrests - Nonindex	12.97 (43.27)	10.78 (38.12)	7.73 (30.1)
Arrests - Drug	3.15 (19.64)	2.78 (18.57)	1.31 (12.58)
Arrests - Quality of Life	6.93 (33.47)	6.2 (30.54)	3.19 (20.37)
Arrests - Simple Violent	1.47 (12.86)	1.46 (12.73)	1.14 (11.06)
Arrests - Weapon	0.21 (4.83)	0.28 (5.52)	0.44 (6.99)
Arrests - Other Nonindex	5.54 (30.13)	3.23 (22.94)	1.82 (15.34)
Arrests - Traffic	0.39 (7.47)	0.89 (9.89)	1.23 (11.27)
Arrests - Municipal	0.69 (9.99)	0.6 (8.86)	0.37 (6.42)
Arrests - Warrant	1.51 (12.66)	1.55 (12.74)	1.43 (12.18)
Arrests - Index	1.61 (14.04)	1.58 (14.06)	1.38 (12.96)
Arrests - Violent Index	0.36 (6.6)	0.4 (7.05)	0.36 (6.37)
Arrests - Property Index	1.25 (12.39)	1.18 (12.16)	1.02 (11.29)
Uses of Force	0.4 (6.51)	0.36 (6.19)	0.29 (5.53)
Uses of Force - Injuries	0.1 (3.19)	0.09 (3.04)	0.04 (2.03)
Complaints	0.25 (5.03)	0.2 (4.5)	0.07 (2.7)
Complaints - Serious	0.17 (4.1)	0.12 (3.49)	0.04 (1.9)
Complaints - Failure to Provide Service	0.05 (2.24)	0.04 (2.06)	0.02 (1.31)
Complaints - Other	0.03 (1.87)	0.04 (1.94)	0.02 (1.36)
Arrests - Index (Primary)	0.83 (9.99)	0.8 (9.89)	0.71 (9.15)
Arrests - Nonindex (Primary)	6.57 (30.44)	5.46 (26.78)	3.96 (21.22)
Arrests - Index (Supp.)	1.22 (13.39)	1.73 (16.14)	2.03 (17.21)
Arrests - Nonindex (Supp.)	9.28 (44.38)	9.87 (42.73)	9.47 (38)

Note: Table presents the means and standard deviations of outcomes across each sample period with each sample filtered to be the analysis sample, meaning the observations belong to the leave-one-out largest connected component. Primary refers to the officer being the first officer, while Supp. refers to them being supplementary or auxilary. By default, the outcomes only count for primary and secondary officers (non-supplementary) when relevant.