What Makes a Good Cop?

Measuring Officer Quality and the Trade-Off Between Policing and Public Safety

Roman Rivera*

McCourt School of Public Policy

Georgetown University

September 22, 2025

Abstract

This paper measures the impact of individual police officers on crime and policing activities (stops, arrests, force, and complaints). I measure officer quality as their effect on total costs, combining the costs of crimes and activities. Officer effects on activities and crime vary significantly, producing large quality heterogeneity. Activity-effects explain a majority of the variation in crime-effects and cost-effects, while officer demographics explain little. Most activities have null effects on crime, though stops do reduce crime costs. Due to direct effects on crime and correlations between activity-effects, high-stop officers reduce total costs, while high-force and high-complaint officers increase total costs.

*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, Aurelie Ouss, Sam Norris, David Card, Steven Raphael, Conrad Miller, Felipe Goncalves, Alessandra Fenizia, 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, Princeton University Industrial Relations Section, and the McCourt School of Public Policy 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). Email: romgariv@gmail.com

1 Introduction

Do police officers differ in quality? A growing literature shows that individual officers vary significantly in their on-the-job activities, including arrests, use of force, and stops and searches (Weisburst (2022), Adger et al. (2022), Goncalves & Mello (2021), Feigenberg & Miller (2022)). However, it is unclear whether these differences in policing activities relate to the primary aim of policing: reducing crime. Furthermore, while prior work shows that police do reduce crime (Di Tella & Schargrodsky (2004), Evans & Owens (2007), Chalfin et al. (2022)), we have not been able to measure whether officers differ in their ability to reduce crime. Difficulty measuring individual performance is common in public-sector professions, where observable activities do not clearly connect to organizational goals (Chetty et al. (2014), Rose et al. (2022), Fenizia (2022), Best et al. (2023), Besley et al. (2022)).

In this paper, I estimate the effects of individual officers on crime reduction and on a range of policing activities (stops, arrests, use of force, and misconduct). Then, I use these officer effects to measure and predict officer quality. I define quality as an officer's impact on total social costs, which factor in the savings from crime reduction and the costs from policing activities (Becker (1968)). This internalizes the trade-off between police reducing crime and the mounting evidence of the negative spillovers from policing activities (Ba (2018), Ouss & Rappaport (2020), Ang et al. (2021)).

Estimating officer effects on activities and crime faces three challenges: officers engage in a wide range of activities; specific crimes cannot be attributed to individual officers; and officers are not randomly assigned to their assignments. I overcome these challenges using detailed data on Chicago Police Department officers between 2010 and 2019, linking granular shift and assignment records with a wide range of policing activities — street and traffic stops, arrests for more serious offenses (violent and property crimes) and for low-level offenses (e.g., drugs, vandalism, weapon possession), use of force, and complaints of misconduct. I use officers' start and end times to construct the team of officers working together within the same location and time, and I link teams to the crimes that occurred while they worked.

Institutional features generate quasi-random variation that identifies officers' causal effects on policing activities and crime. Officers work in various assignments over time, but the police district and time of day they work are relatively stable. Additionally, the exact days an officer works are predictable and determined by a fixed rotation calendar, which also generates variation in the composition of officers working across district-hours. I construct a single metric of officer quality and capture officer effects on costs by applying incidence cost estimates to crimes (Miller et al. (2021)) and conservative cost estimates to each policing activity.

I begin by estimating officers' causal effects on the policing activities at the officer-shift level. I recover unbiased estimates of the (co)variance structure of these officer effects using Kline et al. (2020)'s leave-out method. I find that variation across officers has economically large effects on policing activities, but officers are relatively specialized in their propensities for different activities. For each activity, one standard deviation (SD) in officer fixed effects exceeds 70% of the mean and is precisely estimated. Over the course of a year, a one SD higher officer in each respective activity causes an additional 48 street stops, 36 traffic stops, 17 low-level arrests, 2.6 serious arrests, 0.7 uses of force, and 0.3 complaints. One SD in activity costs effects is over $100,000 per officer-year. This is primarily driven by variation in officer propensities for low-level arrest, which individually impose sizable social costs and are over 20 times more common than force or misconduct. Interestingly, pairwise correlations between officer effects for different activities are positive but modest (median correlation < 0.3), while the largest is between use of force and complaint effects at 0.63.

Next, I show that officers have a modest but economically significant effect on crime and total costs when and where they are working. At the team (district-hour) level, I estimate the effect of individual officers on crimes, policing activities, and costs within that district-hour, controlling for the total number of officers working and employing granular fixed effects that interact district, year-month, day of week, and hour of day. I bias-correct the (co)variance structure using the split-sample method (Silver (2021), Finkelstein et al. (2021)). I find

that officers have economically significant effects on reported crime, as one SD in violent and property crime effects equates to 4 and 12 additional crimes per officer-year (1.8% and 1.4% of their means). For total costs (crime plus activity), one SD in officer effects is $716,000 (1.6%) per officer-year. This is primarily driven by their influence on crime rather than activities, as one SD in officer effects on crime costs amounts to $660,000 (1.7%) per officer-year.

I then use officer demographics and activity effects to explain and predict officer quality. I find that observable officer demographics are not strong predictors of officer quality but officer propensities for policing activities are. I find many statistically significant relationships between officer demographics and activities consistent with prior work (e.g., Ba, Knox, et al. (2021), Hoekstra & Sloan (2022)). However, a battery of officer demographics — race, gender, starting age, tenure, education — only explains between 5% and 17% of variation in activity effects, meaning within-group heterogeneity is large, and between 1.2% and 2.3% of the variation in officer effects on crime and costs. In contrast, the majority of the variation in officer effects on violent and property crime, crime costs, and total costs can be explained by variation in officers' activity effects (50% - 72%). This implies that on-the-job activities are informative about an officer's effect on crime and their overall quality. The dominance of on-the-job activity measures in explaining across-officer variation, relative to demographics, mirrors a common result in personnel and education economics when trying to explain worker, manager, or teacher quality (Jackson et al. (2014), Hoffman & Stanton (2024), Fenizia & Kirchmaier (2025)).

To explore the relationship between officer activities and crime, I combine the precision of the shift-level officer effects with the team-level design. I construct team's collective propensity for each activity by summing over the team members' shift-level officer fixed effects. This allows me to estimate both the 'reduced form' relation between officer propensities for each activity and crime reduction, as well as the causal effect of the marginal policing activity on crime by using team propensities as instruments for the number of policing activities within a district-hour. I find small but precise effects of team propensities and policing activities on

crime. The 2SLS results are robust to tests of the exclusion restriction, suggesting the effect of officer heterogeneity on crime is driven by their influence on activities rather than other unobserved behaviors or demographics.

The elasticities of violent and property crime with respect to all activities are smaller than -0.09, and only two elasticities are statistically significant: property crime with respect to street stops (-0.08) and low-level arrests (-0.06). The street stop and low-level arrest elasticities are contained within confidence intervals of prior elasticity estimates that use higher-level variation (Bacher-Hicks & Campa (2020b), Cho et al. (2023)), and all elasticities are significantly smaller than established elasticities of crime with respect to police quantity (e.g., Chalfin & McCrary (2018), Draca et al. (2011)). Notably, only street stops produce net cost savings due to their significant effects on crime costs ($-$\$2,300 per stop) and their small activity cost, while low-level arrest's crime-cost savings are outweighed by their activity costs. Additional analyses suggest that the effects of activities on crime are highly localized and dissipate within four hours. These results are consistent with the broad literature showing presence and visibility, rather than policing activities, are the primary drivers of crime reduction and that deterrence effects are highly localized in space and time (Chalfin & McCrary (2017), Gonzalez & Komisarow (2020), Weisburd (2021)).

I conclude by determining which officers reduce costs, and thus are higher quality. I first consider targeted replacements of officers based on their propensities for a specific activity, taking into account the cross-activity correlations in officer effects. Replacing an officer who is +1 SD above the mean in use of force or complaints with an average officer generates significant savings of over \$10,000 per officer-year. These savings arise because such officers generate high activity costs (primarily from low-level arrests) do not engage in significantly more street stops, which reduce crime costs. While replacements based on traffic stops or arrests do not generate significant effects, replacing high street stop officers increases costs. This suggests that officers with higher street stop propensities tend to be higher quality, while those with higher force or misconduct propensities tend to be lower quality.

These conclusions depend on how policing activities are valued. If street stops are sufficiently costly (e.g., $500 per stop) due to the social costs of, for example, racial disparities (Fagan & Davies (2000), Gelman et al. (2007)), then I cannot rule out that higher stop officers increase total costs. However, if arrests actually generate social savings due to large incapacitation effects or positive spillovers on individuals' health or families (Norris et al. (2021, 2024)), then high arrest officers and high stop officers both reduce costs relative to the average officer.

By contrast, the targeted selection of officers based on demographics does not produce clear savings, suggesting officer quality is not reliably associated with officer demographics. This is in part because officer demographics explain very small portions of variation in officer quality, and because the demographics associated with lower activity costs cause similarly sized increases in crime costs. This is consistent across alternative activity cost schemes. Overall, officers' on-the-job behaviors are more useful than demographics in selecting higher quality officers.

This paper contributes to the broad literature in labor economics on the importance of heterogeneity across firms, workers, and managers (Abowd et al. (1999), Card et al. (2013), Rothstein (2010), Xu (2018), Bonhomme (2021), Chan & Chen (2022), Petek & Pope (2023)) and the literature on officer heterogeneity in policing activities (e.g., Rozema & Schanzenbach (2019), Reeves (2024). Goncalves et al. (2025)). The central contribution is applying the canonical 'value-added' framework to police officers and measuring the joint distribution of officer effects on multiple observable activities and crime reduction. By applying costs to and aggregating across these dimensions, this paper provides a novel and more holistic measure of officer performance and quality. This complements the literature on how officer quantity influences crime and social costs (Chalfin & McCrary (2017), Owens & Ba (2021)), and also contributes to our understanding of police oversight, training, and diversion programs aimed to reduce costly behaviors (Prendergast (2001), Dube et al. (2025), Davis et al. (2025)). Similar to Goncalves & Mello (2023) and Amaral et al. (2023), I use officer heterogeneity in

policing activities to study the effect of enforcement on offending. Finally, I find evidence of specialization in preferences for different activities, which contributes to the literature on preferences and intrinsic motivation in policing (Prendergast (2007), Chalfin & Goncalves (2023)).[1]

The paper proceeds as follows. Section 2 discusses background and data. Section 3 estimates the effects of officers on activities, crimes, and costs, and predicts officer effects. Section 4 connects activities and crime. Section 5 examines which officers reduce costs. Section 6 concludes.

2 Setting and Data

2.1 Background on the Chicago Police Department

The Chicago Police Department (CPD) is the second largest police department in the United States, policing 2.7 million people with over 10,000 sworn officers. An officer's unit, watch, and day-off group ('DOG') determine where, what time, and on which day they work. In general, each of these is determined at the end of the previous year, and changes occur infrequently through a bidding process based on tenure (Ba, Bayer, et al. (2021)).

Units are the main organizational structure within the CPD. Most officers work in patrol units. Each patrol unit polices a specific police district — one of twenty-two areas of the city. Officers are generally responsible for patrolling, answering 911 calls, and serving on special and tactical teams within their unit's corresponding district. Other units are smaller and have more specialized functions (e.g., training, detectives, gang enforcement, SWAT). The time an officer works their shift is determined by their watch. Most officers work 9-hour shifts on one of three watches, starting around 10 pm, 7 am, and 2 pm. When on duty, an

[1]For effects on crime and social costs see also Levitt (2002), Klick & Tabarrok (2005), Donohue et al. (2013), Weisburst (2019), Mello (2019), Mastrobuoni (2019), Cheng & Long (2018), Cox & Cunningham (2021). For effects of oversight and trainings see Rivera & Ba (2023), Rozema & Schanzenbach (2023), Owens et al. (2018).

officer works a specific beat, which specifies a role or specific location of policing within their unit and watch (e.g., responding to 911 calls in a specific area, bike patrol, desk duty, etc.).[2]

A rotation calendar determines which days an officer works based on their DOG. Between 2010 and 2019, the vast majority of officers worked on a 4-2 rotation (4 days on duty followed by 2 days off duty). The fixed nature of DOGs and the rotation calendar prevents officers from easily or systematically manipulating days on or off duty. Additionally, the exact set of officers working a given unit and watch will change from day to day because of overlapping DOGs (e.g., DOGs A and B work one day while DOGs B and C work the next day), as exploited in prior work (e.g., Ba, Knox, et al. (2021), Gudgeon et al. (2023)).

The calendar can be validated using CPD data to predict days on and off for each officer by taking their first observed 4-days-on/2-days-off cycle within a year and extending the pattern throughout the year. Figure 1 compares officers' predicted rotations with their observed days on and off. Despite ignoring all vacation days, injuries, training, potential for switching units, or changing DOGs, this prediction method is correct for 85% of officer-days early in the year and declines to around 70% correct by the end of the year.[3] For comparison, randomly guessing days on/off correctly predicts $\approx 45\%$ of officer-days.

2.2 Data and Summary Statistics

The primary data for this study encompass the daily activities of CPD officers over ten years (2010-2019), based on daily attendance and assignment records. These data contain information on officer rank, whether they were present for duty, their start and end times, assigned unit, watch, and beat, among other information for each shift. The analysis sample focuses on 'Police Officers', the lowest and most common rank ($\approx 80\%$ of the CPD). Along with a few other restrictions, the analysis sample contains officer-shift observations in which

[2]Unlike units and watches, an officer's assigned beat can change on a day-to-day basis. However, they generally work the same or a small set of beats, and the exact assignment of officers to beats is determined by upper commanders within the unit and watch.

[3]Gudgeon et al. (2023) constructs a similar prediction. In a minority of cases, other rotation patterns are used (5-2 or 6-3), and I impute a 4-2 pattern for the officers for whom no standard pattern exists.

officers are present for duty in active assignments (e.g., not missing, in training, or desk duty). I distinguish between three periods — 2010-2012, 2013-2015, and 2016-2019 — because three small districts were absorbed into larger ones in late 2012, and the scandal surrounding Laquan McDonald's murder by CPD officer Jason Van Dyke began in late 2015. See Appendix C for more details.

I use precise start and end times to construct officer-by-hour data, with most shifts lasting 9 hours. There are two levels of analysis involving similar samples of officers and units constructed from the officer-by-hour data. The first aggregates officer-by-hour data at the officer-by-shift level, and the second aggregates it at the district-by-hour level. Tables 1 and 2 display their respective summary statistics. At the officer-shift level, a single observation corresponds to one officer working on a given day, watch, unit, and beat. Panels A and B of Table 1 show that there are 12,944 unique officers working a total of 11,993,940 officer-shifts, with the average officer working over 1,300 shifts. There are a total of 2,051 unique assignments, defined as the interaction of unit, watch, and beat category. Almost all officers ($> 97\%$) switch between assignments within a sample period, with $> 45\%$ of shifts worked in an officer's most common assignment.

Data on time-stamped officer-level policing activities are merged with assignment data. I distinguish between six activity types that encompass most of the policing activities considered in policy and academic research, summarized in Panel C of Table 1. Street stops (averaging 35 per 900 officer-hours) and traffic stops (13) are the most common activities. Next, I categorize arrests into nonindex (11) and index (1.6) arrests, depending on the associated crime. Index arrests are more serious, either violent or property crimes, generally felonies, and are considered less discretionary. Nonindex arrests are lower-level, more discretionary, and often do not involve a victim. Finally, uses of force (0.36) and complaints of misconduct (0.17) are the least common.[4]

[4]Major reforms in 2016 also resulted in a significant decline in street stops and a rise in traffic stops (Hausman & Kronick (2023)). Index violent crimes: murder, robbery, criminal sexual assault, aggravated assault; index property crimes: burglary, motor vehicle theft, larceny/theft, arson; nonindex crimes include drugs, prostitution, simple battery, vandalism, weapons possession, municipal code violations, traffic offenses,

To quantify the social cost of activities officers engage in, I apply conservative activity cost estimates to each activity, and each should be interpreted as a lower bound (see Appendix C.4 for discussion). Each complaint is $12,000, based solely on the relationship between complaints and settlement costs. A use of force incident is $23,864, equivalent to the cost of a criminal assault excluding crime-related justice costs from Miller et al. (2021). An index arrest is $10,000 and a nonindex arrest is $5,000, based on the costs of pretrial detention relative to release. Street and traffic stops are $0. Over 900 officer-hours, the average officer engages in activities costing $76,900.

The district-hour sample is constructed from police officers working in patrol units ($\approx$ 80% of officer-shifts). I define a 'team' as the set of officers working in the same district and hour (e.g., all Police Officers present for duty in District 12 on June 15, 2013, from 2-3 pm). Table 2 shows that there are 1,835,819 teams and a total of 11,468 unique officers on these teams working in 25 unique districts. The average team contains 49 officers.

I map the crimes reported when a given team is working by summarizing time- and location-stamped crime data (CPD (2025)) to the district-hour level. There are 0.49 index property crimes and 0.13 index violent crimes in the average district-hour. Murder is the rarest crime (0.0026 per district-hour), while theft is relatively common (0.339). (Nonindex) drug crimes are also common (0.123), though they are entirely driven by officer-initiated arrests and cleared by arrest at nearly 100%, unlike other crimes. In contrast, only 9% of property and 17% of violent crimes are cleared (see Table C.4 for clearance rates).

I also map the numbers of stops, arrests, uses of force, and complaints based on the policing data to district-hours based on when and where the activity was reported. Policing activities are counted only once at the district-hour level — at the officer-shift level, activities can be double-counted if multiple officers report being involved in the same incident.

among others. Uses of force range from relatively minor force (e.g., pressure points, defensive techniques) to gunshots and tasers, which are relatively rare. Following the Laquan McDonald scandal at the end of 2015, complaint frequency decreased from approximately 0.24 to 0.1 per 900 officer-hours. 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).

Additionally, all activities are counted if they occurred within the district-hour, even if they were attributed to an officer excluded from the analysis sample (e.g., a detective). In the average district-hour, there are 1.36 street stops, 1.18 traffic stops, 0.44 nonindex arrests, 0.068 index arrests, 0.019 uses of force, and 0.013 complaints, and these activities result in a total activity cost averaging $3,480 per district-hour.

To measure crime costs, I apply incidence cost estimates for each crime type from Miller et al. (2021) (see Table C.1). Because murder is rare and extremely costly (around $8 million each, similar to 200 robberies), I also use a 'low murder' (LM) crime cost, which values murder at $400,000 to avoid increasing imprecision compared with the full murder (FM) results, similar to the procedure in Heckman et al. (2010). FM and LM crime costs average $41,500 and $22,100 per district-hour, and total costs, which sum activity and crime costs, average $45,000 (FM) and $25,600 (LM) per district-hour.

3 Effects of Officers on Policing and Crime

This section employs two complementary empirical strategies to estimate the role of officer heterogeneity on the two components of officer quality, policing activities and crime.

3.1 Empirical Strategies

3.1.1 Shift-Level

I begin by estimating the causal effects of individual officers on their observed policing activities while they are on duty. I then use these officer effects to quantify officer heterogeneity and to study how officer effects covary across activity types. I model officer i's policing activities of type k while working on their shift s in day-hour t ($y^k_{i,s,t}$) as:

$$y^k_{i,s,t} = \theta^k_i + \psi^k_{J(i,s),P} + x'_{i,s,t}\gamma^k + r^k_{i,s,t} \tag{1}$$

where θ^k_i is the officer's fixed effect, or their individual causal effect on the outcome

when they work. $\psi^k_{J(i,s),P}$ is their assignment-by-sample period fixed effect, which interacts assignment (unit $\times$ watch $\times$ beat category) with sample period (2010-2012, 2013-2015, 2016-2019) to account for structural changes in the CPD. $x_{i,s,t}$ contains other time-varying factors (exact date fixed effects in the main specification). All other unobserved factors are contained in $r^k_{i,s,t}$. In practice, I estimate equation (1) at the officer-shift level and divide outcomes by shift duration (e.g., stops per hour on duty).

Identification of the parameters in equation (1) relies on strict exogeneity being satisfied: conditional on officer and assignment effects and observables, all present, past, and future components of $r^k_{i,s,t}$ are all mean zero, $\mathbb{E}[r^k_{i',s',t'}|\theta^k_i, \psi^k_{J,P}, x_{i,s,t}] = 0$. This means that movement of officers between assignments cannot systematically depend on $r^k_{i,s,t}$ (Card et al. (2013)). This is satisfied if, conditional on $\theta^k_i, \psi^k_{J,P}, x_{i,s,t}$, three components of $r^k_{i,s,t}$ are mean zero: 1. match effects between officer i and assignment J in excess of their fixed effects, 2. drift in officer i's fixed effect over different shifts, and 3. shocks to crime and policing conditions in assignment J during shift s observed by the officer.

As discussed in Section 2, officers cannot systematically choose days on duty based on crime and policing conditions due to the rotation calendar, as validated in Figure 1, and they cannot easily switch to different units or watches. I also test for asymmetric match and drift effects using event study plots around assignment switches, following Card et al. (2013)'s event study test. Figures 2-3 display officer activity rates around a switch between assignments, with officers grouped by their assignment's average quantile of that activity (e.g., Q4 of nonindex arrests to Q1 of nonindex arrests). Consistent with exogenous mobility and symmetric match effects, officers moving from higher to lower activity assignments see symmetric changes in activity rates as those moving from correspondingly lower to higher activity assignments. Evidence of drift effects would be officers switching to assignments with more street stops as they gained a taste for them, generating asymmetric pre-trends before the switches. For some outcomes, I do find pre-trends before switches. However, they are inconsistent with drift or match effects because they are parallel across switch types (e.g.,

switches from Q4 to Q4, Q3, Q2, and Q1 in street stops all display similar pre-trends), are small in magnitude, and are driven by the shift right before the switch.

The degree to which across-officer variation influences activities is captured by $Var(\theta_i^k)$, while the degree to which an officer's propensity to engage in one activity k is related to their propensity to engage in another activity l is captured by $Cov(\theta_i^k, \theta_i^l)$. Even with unbiased estimates of officer effects ($\hat{\theta}_i^k$), (co)variances of estimated officer effects will be biased due to sampling error. So, I employ the Kline et al. (2020) (KSS) heteroskedastic-consistent bias-correction to obtain unbiased estimates of the (co)variances and the Lachowska et al. (2023) extension for cross-activity covariances (e.g., $\widehat{Var(\theta_i^k)}, \widehat{Cov(\theta_i^k, \theta_i^l)}$). Standard errors on (co)variance components are computed using nonparametric bootstrap.[5]

3.1.2 Team-Level

Next, I examine the role of police officer heterogeneity on crime as well as activities. As crime is observed in specific locations and times, rather than being attributable to an individual officer, I model the effect of an officer i, on a team ($i \in \tau(d,t)$) working in district d and hour t, on the number of crimes or policing activities of type c occurring ($Y_{d,t}^c$) as:

$$Y_{d,t}^c = \sum_i \omega_i^c \times 1[i \in \tau(d,t)] + \pi^c N_{d,t} + \Gamma_{d,t}^c + \epsilon_{d,t}^c \tag{2}$$

where ω_i^c is the fixed effect of officer i on a given activity or crime c occurring at the district-hour level. Γ^c are interacted district, year, month, day of week, and hour fixed effects ('DYMWH'), and $N_{d,t} = \|\tau(d,t)\|$ is the team size which controls for the total number of officers working. To ensure the design matrix is invertible, I apply the pruning algorithm from Bonhomme (2021).[6] Identification relies on the quasi-random assignment of specific officers

[5]El Karoui & Purdom (2018) show that bootstrap can perform poorly in high-dimensional settings. However, this is for cases where the number of parameters is proportional to the number of observations, whereas in this study the ratio of parameters to observations is less than 0.003. Additionally, they show that as this ratio grows, pair-sampling, as done in this paper, produces more conservative intervals.

[6]I exclude assignment type and exact day fixed effect controls from this analysis due to their minimal influence on explained variation and the increased computational intensity and difficulty ensuring the matrix

to district-day-hours within DYMWH bins. Officers are permitted to select districts, days of the week, months, and hours of day to work, but they are not permitted to choose whether they are working on a specific day-hour based on short-run shocks to policing conditions ($\epsilon^c_{d,t}$) — e.g., whether an officer is working in District 12 at 2pm on the 1st, 2nd, 3rd, or 4th Monday of March 2012 must be unrelated to the shocks to policing conditions within that district-hour. The operations calendar impedes such endogenous selection and also ensures that there is identifying variation in the exactly which officers are working on a specific week day within district-year-month-hour bins due to overlapping DOGs.

The variance of ω^c_i captures the influence of officer heterogeneity on crime and policing activities at the district-hour level, while cross-outcome correlations capture how one's propensity for causing/preventing one type of crime or activity relates to one's propensity for causing/preventing another type of crime or activity. To estimate the variances ($Var(\omega^c_i)$) and cross-outcome covariances ($Cov(\omega^c_i, \omega^{c'}_i)$), I employ the nonparametric split-sample procedure (Silver (2021), Best et al. (2023)). Specifically, I partition the data into two random samples of DYMWHs, and I estimate $\hat{\omega}^c_{i,1}$ and $\hat{\omega}^c_{i,2}$, then compute $\widehat{Var(\omega^c_i)} = Cov(\hat{\omega}^c_{i,1}, \hat{\omega}^c_{i,2})$ and $\widehat{Cov(\omega^c_i, \omega^{c'}_i)} = Cov(\hat{\omega}^c_{i,1}, \hat{\omega}^{c'}_{i,2})$. I repeat this using different random splits forty times, recovering forty unbiased estimates of each (co)variance. I take the mean of these estimates to use as the point estimate and use the standard deviation as the standard error.[7]

This team-level model is complementary to the shift-level model, with both strengths and weaknesses. Its primary strength is that the team model can estimate officer effects on crime and policing activities while accounting for spillovers and avoiding issues with double-counting activities. In contrast, the officer-shift approach properly attributes policing activities to officers while the team model can mis-attribute activities (e.g., each arrest made

is invertible.

[7]While more efficient than the split-sample method, the KSS bias-correction requires finding the leave-one-out connected set, which is straightforward with two-way fixed effect types (e.g., officers and assignments) as in equation (1), but it is not feasible in the case of K-way fixed effects in the case of equation (2). The (co)variance of ω^c_i are computed by transforming the $T \times P$ (number of district-hours $\times$ number of officers) matrix with multiple officer indicators 'on' into a matrix with one row per officer-team and a single officer indicator 'on'.

in a district-hour is attributed equally to each member of the team) and has significantly larger effective observations per officer. As a result, we expect larger sampling error in $\hat{\omega}_i^k$ than $\hat{\theta}_i^k$. Additionally, KSS is more efficient than the split-sample procedure, and so (co)variance components will be less precise when using team-level estimates.

3.2 Officer Heterogeneity in Policing Activities

Panel A of Table 3 presents (co)variance components of shift-level officer effects (θ_i^k) for each of the primary activity types (Columns (1)-(6)) and total activity costs (Column (7)). Officer heterogeneity generates economically significant variation in each activity. Over 900 officer-hours (around half of an officer-year), one standard deviation (SD) increase in officer fixed effects increases street stops by 24.4 (a 71% increase relative to the mean), traffic stops by 17.9 (142%), nonindex arrests by 8.47 (83%), index arrests by 1.26 (83%), uses of force by 0.366 (107%), and complaints by 0.134 (81%). These estimates are precise, with all t-values greater than 20. For comparison, in prior work, 1 SD in complaint effects and use of force effects have larger relative effects (140% of the mean in Rozema & Schanzenbach (2019) and 138% of the mean in Adger et al. (2022), respectively), while Weisburst (2022) finds a smaller 31% increase in arrests per 1 SD increase.[8]

One SD in officer effects on total activity costs is $54,439.95 per 900 officer-hours (72% of the mean). Variation in nonindex arrest propensity is the primary driver of variation in activity costs, as 1 SD in nonindex arrests accounts for 78% of 1 SD in activity costs. By contrast, 1 SD in index arrests accounts for only 23% of 1 SD in activity costs, while 1 SD in use of force and complaints are only 16% and 3%, respectively. This is because nonindex arrests are significantly more common than other costly outcomes but their costs are of a similar order of magnitude — e.g., 1 SD in nonindex arrests is 23 times larger than 1 SD in use of force while use of force is only around 5 times more expensive.

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

For the more frequent outcomes (street and traffic stops and nonindex arrests), between 14% and 22% of $Var(y^k)$ is explained, while for index arrests, complaints, and uses of force, only 0.2% to 2.3% of $Var(y^k)$ is explained. Almost all of this explained variation is due to $Var(\theta_i^k)$ and $Var(\psi_{J,P}^k)$. Notably, very little of the variation in activities is explained by officers sorting into assignments ($2 \times Cov(\theta_i^k, \psi_i^k)$). As a result, there are small correlations between officer and assignment effects (all < 0.2). These patterns suggest modest sorting into assignments based on opportunities for low-level enforcement (e.g., stops, nonindex arrests). However, they do not support a narrative in which high-force or high-misconduct officers sort into roles enabling serious abuse.

Figure A.1 shows that the shift-level estimates are similar to those recovered from the team model. Panel A shows that officer effects on activities are strongly correlated across empirical strategies, as regressing $\hat{\omega}_i^k$ on $\hat{\theta}_i^k$ produces coefficients between 0.26 and 1.1 (all $p < 0.01$), consistent with shift-level design double-counting some activities. Panel B shows that $\widehat{SD(\omega)}$'s are similar to but larger than $\widehat{SD(\theta)}$'s.

I probe the robustness of these results by performing a number of tests discussed in detail in Appendix B.1. To address concerns over endogenous mobility in particular, I show the results are similar using only days when the officer is predicted to be working ("Predicted On"), when the first and last day of an officer's assignment-spell is dropped ("Donut"), or when the data are aggregated to the week-level ("Week") (see Figures A.2-A.3).

3.3 Cross-Activity Correlations

We can determine the degree to which officer propensities for one activity relate to propensities for others based on how officer effects covary across activities. Panel B of Table 3 displays the KSS bias-corrected pairwise correlations between shift-level officer effects for each pair of activities. These correlations are precisely estimated, as all standard errors are < 0.032. Pairwise correlations between activities are not exceptionally large and none exceed 0.7, suggesting officer propensities for different behaviors are not highly interdependent. Notably,

most stop and arrest activities are only modestly correlated with each other. Traffic stops have the weakest cross-outcome correlations with the other stops and arrests (< 0.29). Street stops, nonindex arrests, and index arrests are all modestly correlated with each other between 0.43 and 0.52. The lack of large cross-activity correlations between stop and arrest effects suggests that officers are better characterized as specialists in activities than simply as either high or low activity officers.

The strongest cross-activity correlation is between use of force and complaint effects at 0.63. Complaints are weakly correlated with other activities: the strongest correlation, with nonindex arrests, is only 0.27, followed by weaker correlations with index arrests (0.12), traffic stops (0.11), and street stops (0.067). Use of force effects are more strongly correlated with nonindex arrests (0.38), index arrests (0.33), street stops (0.23), and traffic stops (0.13).[9] Unsurprisingly, nonindex arrests have the strongest correlation with activity costs (0.96), due to both their direct influence on costs and their positive correlations with the other high cost activities, while traffic stops are the least correlated (0.25) as they have no direct influence on costs and have the weakest correlations with other costly activities. Figure A.4 shows that these correlations are robust to a number of alternative specifications and samples. Similarly, Figure A.5 shows that the corresponding correlations using team-level ω_i's are generally similar to but significantly less precise than the correlations using θ_i's.

Overall, officer effects for stops and arrests — commonly viewed as productive policing activities — are not strongly related to officer effects for use of force and complaints, which are both undesirable and costly. These results are inconsistent with a model of misconduct and force being integral components of good policing. Rather, they are more consistent with use of force and especially misconduct being costly behaviors that are not necessarily related to productivity. However, variation in arrests is the primary activity cost driver, suggesting a nuanced relationship between policing and social costs.

[9]The positive correlations suggest that substantial crowd-out or significant multi-tasking constraints are not major concerns.

3.4 Officer Heterogeneity in Crime Reduction

Next, I quantify the importance of officer heterogeneity in crime reduction beyond the role of officer quantity. Officer crime-effects are estimated from equation (2), and (co)variance components are bias-corrected using the split-sample procedure. Table 4 displays results relating to the variation in officer effects on crime (Panel A) and police activities (Panel B). As expected, the bias-correction is impactful as bias accounts for generally over 90% of the plug-in variance in officer effects for crime effects. Nevertheless, with the exception of weapon crime, variation in officers does produce economically significant variation in crimes and activities at the team-level with all other t-values > 4.

For violent and property crime, variation across officers explains 0.004% and 0.007% of variation in reported crime, which is similar to their relative variance for simple violent crime and other nonindex crimes (both 0.006%). These are all 70% smaller than the variation in street stops explained by officers (0.027%). Unsurprisingly, drug crimes are the most influenced by officer heterogeneity, explaining 0.012% of variation, which aligns closely with the share of variation explained in nonindex arrests (0.011%).

Despite relatively small portions of variation explained, officer heterogeneity is economically meaningful. An officer who is 1 SD better at preventing crime (lower ω_i) prevents 2.1 violent crimes (1.8% of the mean), 6.1 property crimes (1.4%), 4.3 simple violent crimes (1.6%), and 5.2 other nonindex crimes (1.4%) over 900 officer-hours. Consistent with the larger influence of officers on discretionary arrests, 1 SD in drug crime effects is relatively large, equivalent to 'preventing' 4.4 drug crimes (4.1%). Similarly, SDs in officer effects on policing activities are also larger, ranging from 2.3% to 8.1% of their means.

To validate these effects on crime, I use 911 calls as a measure of crime reports unrelated to police officer reporting. Figure A.7 shows that officer effects for violent and property crime are correlated with those for various call types in expected ways: officers who reduce violent crime also reduce calls related to robbery, battery, and general violence, while officers who reduce property crime also reduce calls about theft and burglary. Additionally, officer effects

explain similar portions of the variation in call frequency as they do in crime reports. This suggests the crime effect results are due to differences in actual crime prevention rather than being driven by differential reporting or suppression.

Panel C of Table 4 displays results for activity costs, crime costs, and total costs using the low murder costs — using full murder costs produces estimates that are too noisy to reliably detect officer effects as 100% plug-in bias cannot be rejected. Officers explain 0.01% of the variation in activity costs, but only 0.004% of the variation in crime costs, and thus explain 0.005% of total cost variation. Over 900 officer-hours, this means a 1 SD lower activity cost officers saves $71,000 (2.3% of the mean), a 1 SD lower crime cost officer saves $333,000 (1.7%), and a 1 SD lower total cost officer saves $358,000 (1.6%). Overall, variation in officer quality, as measured by their influence on costs, is large and economically significant. Importantly, the majority of officer heterogeneity in total costs is driven by their influence on crime costs, not activity costs.

3.5 Decomposing Officer Heterogeneity

Given the meaningful variation in officer effects on policing activities, crime reduction, and costs, I consider two sets of variables to explain the variation in officer effects: officer demographics and officer effects on activities. First, I consider the role of a battery of officer demographics: race, gender, tenure, start age, education, and Spanish speaking ability. These features are also appealing because they are easily observable and can be used for recruitment policy, and they are often the focus of policy debates over police composition. I project officer effects $(\hat{\theta}_i^l, \hat{\omega}_i^c)$ onto officer demographics (z_i, indexed by v):

$$\begin{aligned} \widehat{\theta}_i^l &= \sum_v \alpha_v^l \times z_i^v + \epsilon_i^l \\ \widehat{\omega}_i^c &= \sum_v \alpha_v^c \times z_i^v + \epsilon_i^c \end{aligned} \tag{3}$$

I measure the explanatory power of officer observables from the resulting R^2. To account for the measurement error in the dependent variables, I construct a Corrected $R^2 = R^2 \frac{Var(\hat{\theta}_i^l)}{\widehat{Var(\theta_i^l)}}$

which accounts for the amount of explainable variation in officer effects.

Table 7 displays the relationships between officer demographics and officer effects on activities ($\hat{\theta}_i^l$) (Columns (1)-(6)) and officer effects on crime and total costs ($\hat{\omega}_i^c$) (Columns (7)-(10)). The relationships between officer demographics and activities are highly statistically significant and consistent with prior work.[10] Black, female, older, and more tenured officers make fewer stops, arrests, uses of force, and complaints than their white, male, younger, and low-tenured counterparts. Activity levels decline most substantially with tenure: officers with > 20 years of tenure make almost 40% fewer index arrests relative to the mean compared with new officers (< 5 years), while the Black-White or Female-Male gaps are both under 15%. Officer demographics explain a small portion of the variation in officer activity effects, between 5% and 17% after adjusting for bias in $Var(\hat{\theta}_i^l)$, but this indicates that within demographic variation is substantial relative to the average differences between groups. There are few statistically significant estimates for the relationship between officer demographics and team-level officer effects on crime and costs (Columns (7)-(10)), and officer demographics explain only 2.2% of the variation in officer effects for property crime and 1.2-1.5% of the variation in officer effects for violent crime, crime costs, and total costs.

Next, I explore the explanatory power of officer activity effects, which are not directly observable but can be measured. The ideal regression would be regressing officer effects for one outcome (e.g., a specific activity l or crime or cost outcome c) on a vector of officer activity effects $\forall k \in \{1, ..., K\}$, excluding l if $l \in \{1, ..., K\}$:

$$\begin{aligned} \omega_i^c &= \sum_k \beta_k^c \times \omega_i^k + \epsilon_i^c \\ \theta_i^l &= \sum_{k \neq l} \beta_k^l \times \theta_i^k + \epsilon_i^l \end{aligned} \tag{4}$$

Estimating this model with OLS using $\hat{\omega}_i$'s and $\hat{\theta}_i$'s is not feasible because of sampling error in both dependent and independent variables. However, the implied regression estimates

[10] Prior work shows that officer race, gender, age, and tenure are strongly related to their stops, arrests, and use of force. See McCrary (2007), Miller & Segal (2018), Ba, Knox, et al. (2021), Goncalves & Mello (2021), Ba, Bayer, et al. (2021), Hoekstra & Sloan (2022), Cox et al. (2021), Weisburst (2022), Rivera (2025).

of β_k^c, β_k^l can be recovered using the bias-corrected (co)variances (Rose et al. (2022)).[11]

The implied regression results of regressing one activity effect on all others are displayed in Panel A of Figure 4, with coefficients scaled to be conditional correlations ($\hat{\beta}_k^l \times \frac{\widehat{SD(\theta_i^k)}}{\widehat{SD(\theta_i^l)}}$) along with implied R^2's, indicating the share of $\widehat{Var(\theta_i^l)}$ that can be explained by its relationship with all other activity effects. Notably, the pairwise correlation between use of force and complaints is robust to controlling for other effects, while most other relationships become significantly smaller than their pairwise correlations. The main takeaway is that relatively small portions of the variation in one activity effect is explained by all other activity effects, with all $R^2 < 0.5$.[12]

Panel B of Figure 4 displays the implied regression coefficients and R^2's from estimating equation (4) with officer crime and cost effects as the outcome and all officer activity effects as the explanatory variables. Estimates are transformed into conditional correlations ($\hat{\beta}_k \times \frac{\widehat{SD(\omega_i^k)}}{\widehat{SD(\omega_i^c)}}$). The imprecision in the (co)variances of $\hat{\omega}_i$'s results in noisy point estimates, with only one conditional correlation's 95% confidence interval rejecting 0 — officers who increase index arrests also increase property crime reports. Interestingly, conditional correlations with crime costs and total costs are relatively small in magnitude (all but one are < 0.22). Additionally, only street stops have negative conditional correlations with all outcomes — violent crime (-0.08), property crime (-0.02), crime costs (-0.18), and total costs (-0.21).

Despite this imprecision, variation in policing activities explains the majority of each outcome, with R^2 values of 0.57 (violent crime), 0.72 (property crime), 0.5 (crime costs), and 0.52 (total costs). Though estimates are too imprecise to make clear inferences, these results suggest that the influence of officer heterogeneity on policing activities and their effects on crime and costs are strongly connected. Despite using only six activity types, the explanatory power of officer activity effects is larger than that of even a rich set of officer demographics.

[11] For the first line in equation (4) $[\hat{\beta}_1^c, ...\hat{\beta}_K^c] = (\hat{\Sigma}_K' \hat{\Sigma}_K)^{-1} \hat{\Sigma}_K' \hat{\Sigma}_l$, where $\hat{\Sigma}_K$ is the $K \times K$ bias-corrected variance-covariance matrix for ω_i^l's, and $\hat{\Sigma}_k$ is an $K \times 1$ vector of covariances of $\widehat{Cov(\omega_i^k, \omega_i^c)}$.

[12] Figure A.6 shows the same figure using team-level results, which are similar to the main shift-level results but very imprecise by comparison.

The limited ability of officer demographics to explain variation in officer activity, crime, and cost effects is consistent with existing work in education and personnel economics which find that worker observables explain relatively small portions of worker performance (Rockoff et al. (2011), Jackson et al. (2014), Hoffman & Stanton (2024), Weidmann et al. (2024), Fenizia & Kirchmaier (2025)).

4 Effect of Policing on Crime and Costs

This section documents how crime is influenced by officer propensities for policing activities and the direct effects of policing activities on crime.

4.1 Empirical Strategies

The empirical strategies of this section leverage the precision of the shift-level estimates of officer activity effects to more precisely measure the relationship between officer heterogeneity in activities and crime by combining the strengths of equations (1) and (2). I model the 'reduced form' effect of officer heterogeneity in policing activities on crimes of type c in district-hour d, t as:

$$\mathrm{Y}^c_{d,t} = \sum_k \lambda^c_k \Theta^k_{d,t} + \pi^c N_{d,t} + \sum_k \nu^c_k \Psi^k_{d,t} + X'_{d,t}\gamma^c + \Gamma^c_{d,t} + \varepsilon^c_{d,t} \tag{5}$$

where $\Theta^k_{d,t} = \sum_{i \in \tau(d,t)} \hat{\theta}^k_i$ is the summation of officer fixed effects for each activity (recovered from equation (1)) across officers working in that team / district-hour, $\Psi^k_{d,t} = \sum_{i \in \tau(d,t)} \hat{\psi}^k_{J(i,s)}$ is the summation of assignment effects (recovered from equation (1)) worked by team $\tau(d,t)$, and $X_{d,t}$ contains district-hour observables (e.g., exact day fixed effects), and $\Gamma^c_{d,t}$ are DYMWH fixed effects. In practice, $\hat{\theta}^k_i, \hat{\psi}^k_j$ are estimated within each sample period (2010-2012, 2013-2015, 2016-2019) to better account for changes in outcomes when constructing $\Theta^k_{d,t}, \Psi^k_{d,t}$. λ^c_k captures the effect on crimes of type c of having a team which polices more intensely in dimensions k, specifically the effect of increasing a team's propensity

for activity k by 1 (e.g., increasing a single θ_i^k by 1 for some $i \in \tau(d,t)$).

Identification of λ_k^c relies on a strict exogeneity assumption, which is satisfied if $\Theta_{d,t}^k$ is conditionally unrelated to $\epsilon_{d,t}^c \forall t$. As before, institutional constraints and the rotation calendar largely prevent officers from selecting the time of day they work or easily switching between the unit and watch they are assigned to on short notice, particularly when conditioning on DYMWHs.

With additional assumptions, specifically relevance and exclusion, we can recover the direct effect of the marginal policing activity on crime by estimating the two-stage least squares (2SLS) model with the following second and first stages and instrumenting for the number of policing activities of type k using team propensities:

$$\mathrm{Y}_{d,t}^c = \sum_k \eta_k^c Y_{d,t}^k + \pi^c N_{d,t} + \sum_k \nu_k^c \Psi_{d,t}^k + X_{d,t}' \gamma^c + \Gamma_{d,t} + \varepsilon_{d,t}^c \tag{6}$$

$$Y_{d,t}^k = \rho_k^k \Theta_{d,t}^k + \sum_{l \neq k} \rho_l^k \Theta_{d,t}^l + \nu_1^k N_{d,t} + \nu_k^k \Psi_{d,t}^k + \sum_{l \neq k} \nu_l^k \Psi_{d,t}^l + \nu_2^k X_{d,t} + \Gamma_{d,t}^k + \epsilon_{d,t}^k \tag{7}$$

Identification of η_k^c relies on the strict exogeneity assumption holding, and it requires two additional IV assumptions: exclusion and relevance. Relevance — that team activity propensities do influence police activities — can be shown empirically. The exclusion restriction requires that the only mechanism through which activity propensities influence crime is through their influence on policing activities.

4.2 Effect of Policing on Crime

Columns (1) and (2) of Table 5 display the reduced form effects of team activity propensities on violent and property crimes (equation (5)). The coefficients are generally negative and small in magnitude. Only propensities for street stops, traffic stops, and nonindex arrests have (marginally) significant effects on property crime, and no team propensity has a significant effect on violent crime. Increasing a team's propensity to make street stops by 1 street stop per hour decreases property crime by -0.01 ($p < 0.01$), while the effects of traffic stop and

nonindex arrest propensities on property crime are -0.009 ($p < 0.1$) and -0.02 ($p < 0.1$). Teams more prone to complaints and use of force have no significant effects on either violent or property crime. Robustness tests are shown in Figure A.8 and discussed in Appendix B.2.

Next, we turn to the marginal effect of a policing activity using the 2SLS design (equations (6) and (7)). Columns (3) and (4) of Table 5 show that the team activity propensities are strong instruments for observed activities and satisfy relevance. Column (3) displays the first stage coefficient between the focal type of policing intensity ($\hat{\rho}_k^k$ in equation (7)), all of which are economically and statistically significant. Column (4) displays the Sanderson & Windmeijer (2016) F-statistics used to assess weak instruments with multiple endogenous variables, which are large (> 90) for all activities other than complaints and index arrests (> 30). The strong relationships between team propensities and observed activities suggest that the additive construction of $\Theta_{d,t}^k$ is not seriously misspecifying a highly nonlinear relationship between officers and team behavior.

Columns (5) and (6) display the corresponding 2SLS estimates ($\hat{\eta}_k^c$), and Columns (7) and (8) transform these 2SLS estimates into elasticities by dividing by the mean crime rate and multiplying by the mean activity rate ($\hat{\eta}_k^c \times \frac{\overline{Y_{d,t}^k}}{\overline{Y_{d,t}^c}}$) for ease of interpretation and comparability with existing work. The interpretation of these estimates depends on whether the exclusion restriction holds. As discussed in Appendix B.2, I find these effects are consistent across a number of alternative specifications, such as controlling for team demographics and measures of other activity levels. This suggests that exclusion violations are not driving results and that officer heterogeneity is influencing crime through activities rather than behaviors and demographics correlated with those activities.

When transformed into elasticities, effects on violent and property crime are all quite small and precise, with no policing activity having a significant effect on violent crime and only street stops and nonindex arrests having statistically significant negative effects on property crime. By comparison, the OLS estimates of equation (6) in Columns (9) and (10) tend to be biased upward, particularly for the effect of arrests, consistent with crime and

policing being by highly localized shocks to crime conditions in $\varepsilon_{d,t}$.

For violent crime, all elasticities are more positive than -0.09 with standard errors smaller than 0.08. The elasticities of property crime with respect to street stops and nonindex arrests are small, negative, and statistically significant, -0.082 ($p < 0.01$) and -0.06 ($p < 0.05$), while elasticities for all other activities are more positive than -0.025 and not statistically significant with all standard errors smaller than 0.045.[13] Figure A.9 displays a number of robustness specifications as discussed in Appendix B.2.

Figure 5 highlights that the elasticity estimates are small and precise compared with prior works' estimates of elasticities of crime with respect to arrests (Cho et al. (2023)), stops (Bacher-Hicks & Campa (2020b)), and police quantity (Chalfin & McCrary (2018), Draca et al. (2011)). Cho et al. (2023) use officer deaths as an instrument for departmental arrests, and their estimated elasticities are 0.42 for violent and -0.15 for property crime but both are imprecise. Bacher-Hicks & Campa (2020b) use precinct commander variation in New York City to instrument for stops and find positive but imprecise elasticities of felony crimes. While the elasticities using officer-level variation in this paper are smaller and more precise, they fall within the others' confidence intervals.

Next, the elasticities with respect to activities are small in comparison with existing elasticities with respect to officer quantity. Chalfin & McCrary (2018) estimate elasticities of crime with respect to police quantity of -0.34 for violent crime and -0.174 for property crime at the department-by-year level. Draca et al. (2011) find elasticities of selected set of crimes (violent crimes, robberies, and thefts) with respect to short-run manpower within a city at approximately -0.38. The elasticities in this paper are sufficiently precise to reject elasticities

[13]Table A.1 provides elasticities for all index and nonindex crime types. Elasticities for nonindex crimes, which are more determined by officer discretion, are large and significant. For example: drug crimes are highly responsive to nonindex arrests ($+1.3$ ($p < 0.01$)) and also complaints ($+0.14$ ($p < 0.01$)); weapons offenses (cleared over 70%) have positive, large, and statistically significant elasticities with respect to street stops, index arrests, uses of force, and complaints (all > 0.2 and $p < 0.1$), suggesting that these activities lead to reported discovery of weapon possession; and simple violent crimes, such as assault on an officer and domestic violence, have a small positive and relatively precise elasticity with respect to use of force, $+0.046$ ($p < 0.05$), consistent with officers using force causing additional crimes (resisting arrest, assaulting an officer) and resistance by suspects that could otherwise be avoided (Dube et al. (2025)).

of these magnitudes across all policing activities, suggesting a larger role for police quantity than policing intensity in reducing crime. This is consistent with the literature documenting the importance of localized deterrence and visibility rather than policing actions themselves (Gonzalez & Komisarow (2020), Chalfin, Hansen, Lerner, et al. (2022)) and the broader criminal deterrence literature (Chalfin & McCrary (2017)).

4.3 Effect of Policing on Costs

Table 6 displays the effects of officer heterogeneity and policing activities on activity costs, LM crime costs, and LM total costs. Columns (1)-(3) provide reduced form estimates of regressing costs per district-hour on team propensities using equation (5) (see Table A.2 for FM estimates). Columns (4)-(6) provide the effects of the marginal policing activity using the 2SLS approach (equations (7) and (6)); Columns (7)-(9) provide the implied regression coefficients of regressing ω_i^c on $\{\omega_i^k\}_K$ for activities (equation (4)). Reduced form and 2SLS estimates of activity costs closely mirror their assigned values. And while the implied regression coefficients are very imprecise, they generally align directionally with the reduced form and 2SLS results.

Only street stops and nonindex arrests have negative effects on crime costs, and only street stops have a statistically significant effect. One additional stop per district-hour prevents \$2,350 ($p < 0.01$) in crime costs. The reduced form coefficient is -\$910 ($p < 0.01$), meaning an officer with a propensity for 0.05 additional street stop per hour ($\approx 2 \times \widehat{SD(\theta_i^k)}$), but with the same propensity for all other activity effects, reduces crime costs by \$45 per officer-hour. The implied regression coefficient for street stops are similar to the reduced form and 2SLS estimates (-\$1,540 for crime costs and -\$1,900 for total costs, both $p > 0.1$).

One nonindex arrest reduces crime costs by -\$3,400 ($p > 0.1$) and the reduced form coefficient is -\$940 ($p > 0.1$). However, nonindex arrests have a small positive effect on total costs due to their relatively large effect on activity costs and modest effect on crime costs, though if their activity cost were significantly lower they would reduce total costs. Traffic

stops and index arrests have positive and small effects on crime costs, while both complaints and use of force increase crime costs and are very imprecise. As a result, only street stops reduce total costs, and the magnitude is sufficiently large that even a higher estimate for the social cost per stop (e.g., $500) would be significantly smaller than its effect on crime.

4.4 Interpretation and Discussion

The main results implicitly assume the effect of policing on crime is highly localized. If this is not the case due to temporal spillovers, feedback loops, or long-run deterrent effects, then the main estimates will not capture the full effects of activities on crime. For example, an arrest may increase perceived probabilities of detection for would-be offenders in the future or change staffing decisions in the short-run.

I test for this in two ways. First, I estimate the effect of activities in the present (t) on crime in the future ($t + h$) within the same district, which captures persistence in effects. Second, I aggregate the district-hour data to larger time intervals (district by 2-, 4-, 8-, and 24-hour periods), which captures persistence in team composition, attribution of activities to specific time periods, and short-run feedback loops between crime reduction and future staffing. The results are displayed in Figure A.10 and A.11. Consistent with highly localized effects and that the main specification is not misspecified, I find no evidence of long-run persistence beyond 1 hour for nonindex arrests and 4 hours for street stops, and I find very similar elasticities across every level of aggregation.[14]

These small and highly-localized effects are consistent with the existing empirical literature in economics of crime, and it is unlikely these estimates are significantly understating the effects of policing activities. First, the magnitudes of this paper's micro-elasticities are entirely consistent with those using larger-scale variation to study the role of policing activities on crime. For example, larger scale stop declines, arrest reductions, and declines in clearance

[14]Notably, controlling for future team composition ($\hat{\Theta}^{k}_{d,t+h}$) nullifies most of the significant estimates. This suggests that even short-run impacts likely reflect persistence in team composition rather than causal spillovers.

rates generally have null effects on crime (Bacher-Hicks & Campa (2020b), Tebes & Fagan (2025), Cho et al. (2023), McCrary (2007)), while oversight-induced reductions in policing intensity only coincide with increases in crime when there is also a scandal (Rivera & Ba (2023), Devi & Fryer (2020), Premkumar (2022)). While these effects do not speak to effects of mass police pullbacks or going from zero enforcement to some enforcement, it does suggest the marginal and average policing activity are similarly effective.[15] Similarly, spillover effects are small, as Dominguez (2025) shows that gang sweeps have small but detectable deterrence effects on first peers of the arrested offender but not second peers, reinforcing a lack of large scale deterrence effects from an arrest.

The empirical literature on crime reduction is also consistent with would-be offenders responding to contemporaneous police presence (Cohen & Ludwig (2003), Weisburd (2021)), with little evidence of long-run or reputational effects, with the exception of domestic violence arrests (Amaral et al. (2023)). This is in part due to offenders being present-oriented (Lee & McCrary (2016)) with a weak relationship between actual and perceived risk of detection (Lochner (2007), Anwar & Loughran (2011), Apel (2013)). Similarly, crime elasticities with respect to officer quantity are similar regardless of the level of variation, from beat-by-hour to department-by-year (Weisburd (2021), Draca et al. (2011), Cheng & Long (2018), Chalfin & McCrary (2018)), consistent with deterrence effects being highly localized in space and time.[16]

[15] As shown in Table C.5, the 86% decline in street stops between 2015 and 2016 coincided with only a 8.2% increase in property crime, consistent with an elasticity of -0.095 even ignoring the major policing scandals and community outrage that were occurring in that year.

[16] For longer-run elasticities, see also Evans & Owens (2007), Lin (2009), Chalfin & McCrary (2018), Mello (2019), Weisburst (2019); for short-run, see also Di Tella & Schargrodsky (2004), Klick & Tabarrok (2005), Donohue et al. (2013). Facchetti (2025) shows crime increases following station closures are largest near the closed station. For non-police deterrence effects see Gonzalez & Komisarow (2020) and Chalfin, Hansen, Lerner, et al. (2022). See Chalfin & McCrary (2017) for a review.

5 Policy Counterfactuals

This section explores how total costs change under two sets of policies: replacing officers based on their activity effects with average officers and recruiting/retaining officers based on their demographics. I include robustness tests using four alternative activity cost weightings: "Arrests Reduce Costs", nonindex arrests and index arrests generate *negative* activity costs of -$2,500 and -$5,000; "Low Cost Arrests", nonindex arrests and index arrests generate smaller activity costs of $2,000 and $0; "High Cost Force and Complaints", use of force and complaints generate higher costs of $100,000 and $60,000; and "High Cost Stops", street and traffic stops generate high costs of $500 each. While these counterfactuals focus on replacing officers, holding quantity fixed, Appendix B.4 explores the effects of officer quantity on costs.

5.1 Replacing a +1 SD Officer with the Average Officer

First, I quantify the change in activity, crime, and total costs from replacing an officer who is one SD above the mean in each activity type (θ_i^k) with an average officer.[17] I apply the reduced form estimates (equation (5)) to the standard deviations of officer effects for each activity and take into account cross-activity covariances: $\hat{\Delta}^c_{\text{-1SD in } \theta^k} = -1 \times \widehat{SD(\theta_i^k)} \times \left(\hat{\lambda}_k^c + \sum_{l \neq k} \widehat{Cov(\theta_i^k, \theta_i^l)} \times \hat{\lambda}_l^c\right)$. Figure 6 displays the results using low murder estimates per 9-hour officer-shift (see Table A.3 for all estimates). For robustness, it also includes 'Correct Sign Coefs.' estimates which exclude reduced form coefficients with unexpected directions — i.e., no activity can increase crime costs and no activity can decrease activity costs.

I find that targeting officers based on their activity effects can reduce total costs only if high use of force and complaint officers are replaced. The 1 SD decline in use of force and complaint effects saves $131 ($p < 0.01$) and $100 ($p < 0.01$) in activity costs per officer-shift. Note that the savings are driven more by the reduction in nonindex arrests than by the reduction in use of force and complaints. The 1 SD declines increase crime costs by only $38

[17] Chalfin & Kaplan (2021) considers a related counterfactual focusing on how many complaints are prevented by replacing high complaint officers with median complaint officers.

and \$4 (both $p > 0.1$), and the larger rise in crime costs due to force declines is a result of the stronger correlations with nonindex arrests and street stops. On net, the 1 SD declines save \$93 ($p = 0.12$) and \$96 ($p < 0.05$) per officer-shift, indicating that high force and complaint effects are signals of low quality. Using the correct sign coefficients slightly reduces total savings to \$55 and \$69. As a result, replacing a +1 SD officer in force or complaints with an average officer saves at least \$10,000 in total costs per officer-year (around 200 shifts).

In contrast, replacing higher nonindex arrest officers yields the largest activity cost savings (-\$164 ($p < 0.01$)) but also large crime cost increases (+\$156 ($p < 0.05$)), resulting in no significant total cost savings — index arrests and traffic stops have similar but smaller effects. Replacing high street stop officers yields some savings in activity costs (-\$74 ($p < 0.01$)), largely due to correlations with nonindex arrests, but causes significant increases in crime costs (+\$229 ($p < 0.01$)), and results in significant total cost increases (+\$155 ($p < 0.01$)) — which suggests that high street stop officers are higher quality.

Figure A.12 compares effects on total costs across alternative activity cost constructions, with three main takeaways. First, savings from replacing high use of force and complaint officers depends on how costly low-level arrests are. Second, if street stops are sufficiently costly (e.g., \$500 per stop), then I cannot rule out that higher stop officers increase total costs. This would be the case if stops have significant negative spillovers due to racial profiling, educational losses, or damaged community trust (Fagan & Davies (2000), Gelman et al. (2007), Bacher-Hicks & Campa (2020a), Bell (2017)). Finally, if arrests actually have large incapacitation effects or positive spillovers on individuals' health or families (Norris et al. (2021, 2024)), and thus generate negative activity costs, then both higher stop and high arrest officers reduce total costs compared with the average officer.

5.2 Officer Demographics

Next, I estimate the effects of demographics on costs using three methods. First, I estimate equation (8) by regressing costs on team demographics, effectively replacing $\Theta_{d,t}$ in equation

(5) with team demographics ('Direct'). Second, I use estimates from projecting officer effects onto officer demographics (α_v^c in equation (3)) ('Projection'). Finally, I multiply the projection coefficients of demographics on activity effects by the reduced form estimates of activity effects on costs from equations (3) and (5), as shown in equation (9) ('Indirect').

$$\mathrm{Y}_{d,t}^c = \sum_v \zeta_v^c \Big(\sum_{i \in \tau(d,t)} z_i^v \Big) + X_{d,t}' \gamma^c + \Gamma_{d,t}^c + \varepsilon_{d,t}^c \tag{8}$$

$$\iota_v^c = \hat{\lambda}_l \times \hat{\alpha}_v^l \tag{9}$$

Figure 7 displays the results for the effect of each officer demographic on costs, interpreted as the effect of having an officer of a specific demographic instead of the reference officer on duty for a 9-hour shift, and Table A.4 displays effects including full murder costs.[18] Estimates are noisy and generally not statistically distinguishable across estimation methods. The similarity between 'Indirect' method results and those that directly estimate the relationship between crime and demographics is consistent with officer effects on crime flowing through their influence on activities. Demographics associated with lower activity rates generate fewer activity costs but also higher crime costs. I cannot reject such groups produce no cost savings, and some point estimates suggest total cost increases.[19] Figure A.13 shows that using alternative activity cost constructions does not change this general result. In comparison with targeting officers based on their activity effects, the results are consistent with the relatively low explanatory power of officer demographics and suggest that officer quality cannot be reliably predicted using officer demographics.

6 Conclusion

Police officers are street-level bureaucrats with significant discretion over many high-stakes decisions. Their main objectives, to serve communities and to protect them from crime, can

[18]Table A.5 displays effect of demographics on crimes using the direct method (OLS) and an IV specification in which actual team composition is instrumented for using predicted composition based on the calendar.

[19]For example, 'Direct' method estimates suggest that high tenure officers (> 20 years) increase crime costs by over $500 in excess of their activity cost savings, suggesting they are significantly worse at preventing crime than less experienced officers which is inconsistent with Ba, Bayer, et al. (2021).

be at odds with each other as policing activities can generate large costs, making overall quality difficult to measure. This paper leverages detailed data on Chicago police officers to study officer quality in terms of officer heterogeneity in the main policing activities (stops, arrests, use of force, and misconduct), crime reduction, and social costs. I find that officer heterogeneity plays a significant role in determining the number of stops, arrests, uses of force, and complaints of misconduct generated during an officer's shift, though these dimensions are only modestly correlated with each other. As a result, there is significant heterogeneity in activity costs generated by officers, primarily driven by variation in low-level arrests.

Officer heterogeneity also generates economically significant variation in crime and total costs, meaning officer quality varies greatly. This variation in officer effects on crime-reduction and total costs is mostly explained by officer activity effects but not by officer demographics. I estimate effects of policing activities on crime and costs, and I find small but precisely estimated elasticities of crime with respect to activities. Only street stops have economically and statistically significant effects on total costs. As a result of their null effects on crime and their weak relationship with street stops, replacing officers who use force or commit misconduct at high rates can generate significant cost savings.

Policing and crime reduction are complex and nuanced topics. This paper provides an imperfect but tractable definition of officer quality, and it also dispels some assumptions of policing and raises new questions. In particular, I find no evidence that the officers who generate high social costs through misconduct and force are also best at reducing crime or more productive. While propensities for the six focal activities do explain a majority of variation in officer effects on costs, there is still significant variation to be explained that is not attributable to officer activity effects or demographics. Furthermore, the cost calculations relied on a number of conservative lower bounds, meaning more precise metrics of officer quality can be computed once the costs of policing activities are better understood. Uncovering features and behaviors predictive of officer quality and constructing more holistic measures of officer quality and activity costs are ground for future research.

7 References

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

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

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

Amaral, S., Dahl, G. B., Endl-Geyer, V., Hener, T., & Rainer, H. (2023). *Deterrence or backlash? Arrests and the dynamics of domestic violence* (30855). National Bureau of Economic Research. https://www.nber.org/papers/w30855

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2021). *Police violence reduces civilian cooperation and engagement with law enforcement* [Working Paper].

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2025). Community engagement with law enforcement after high-profile acts of police violence. *American Economic Review: Insights*, *7*(1), 124–142.

Anwar, S., & Loughran, T. A. (2011). Testing a bayesian learning theory of deterrence among serious juvenile offenders. *Criminology*, *49*(3), 667–698.

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *Journal of Quantitative Criminology*, *29*(1), 67–101.

Ba, B. (2018). *Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in chicago* [Working Paper].

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.

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

Bacher-Hicks, A., & Campa, E. de la. (2020b). *The impact of new york city's stop and frisk program on crime: The case of police commanders*.

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

Bell, M. C. (2017). Police reform and the dismantling of legal estrangement. *The Yale Law Journal*, *126*(7), 2054–2150.

Besley, T., Burgess, R., Khan, A., & Xu, G. (2022). Bureaucracy and development. *Annual Review of Economics*, *14*(1), 397–424.

Best, M. C., Hjort, J., & Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review*, *113*(8), 2121–2167.

Bhuller, M., & Sigstad, H. (2024). 2SLS with multiple treatments. *Journal of Econometrics*.

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

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., & Goncalves, F. M. (2023). *Professional motivations in the public sector: Evidence from police officers* (31985). National Bureau of Economic Research. https://www.nber.org/papers/w31985

Chalfin, A., Hansen, B., Lerner, J., & Parker, L. (2022). Reducing crime through environmental design: Evidence from a randomized experiment of street lighting in new york city. *Journal of Quantitative Criminology*, *38*(1), 127–157.

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., & Kaplan, J. (2021). How many complaints against police officers can be abated by incapacitating a few "bad apples?" *Criminology & Public Policy*, *20*(2), 351–370.

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, J., David C., & Chen, Y. (2022). *The productivity of professions: Evidence from the emergency department* (30608). National Bureau of Economic Research. https://www.nber.org/papers/w30608

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 II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679.

Cho, S., Goncalves, F., & Weisburst, E. (2023). *The impact of fear on police behavior and public safety* (31392). National Bureau of Economic Research.

Cohen, J., & Ludwig, J. (2003). Policing crime guns. In J. Ludwig & P. J. Cook (Eds.), *Evaluating gun policy* (pp. 217–250). Brookings Institution Press.

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.

Cox, R., Cunningham, J. P., & Ortega, A. (2021). *The impact of affirmative action litigation on police killings of civilians* (p. 84) [Working Paper].

CPD. (2025). *Crimes - 2001 to present |city of chicago |data portal*. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data

Davis, J., Norris, S., Schmitt, J., Shem-Tov, Y., & Strickland, C. (2025). *Mobile crisis response teams support better policing: Evidence from CAHOOTS* (33761). National Bureau of Economic Research. https://www.nber.org/papers/w33761

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.

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.

Ding, P. (2021). The frisch–waugh–lovell theorem for standard errors. *Statistics & Probability Letters*, *168*, 108945.

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.

Dominguez, M. (2025). Sweeping up gangs: The effects of tough-on-crime policies from a network approach. *American Economic Journal: Economic Policy*.

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. (2025). A cognitive view of policing. *The Quarterly Journal of Economics*, *140*(1), 745–791.

El Karoui, N., & Purdom, E. (2018). Can we trust the bootstrap in high-dimensions? The case of linear models. *Journal of Machine Learning Research*, *19*(5), 1–66.

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

Facchetti, E. (2025). *Police infrastructure, police performance, and crime: Evidence from austerity cuts* [Working Paper]. Working Paper.

Fagan, J., & Davies, G. (2000). Street stops and broken windows: Terry, race and disorder in new york city. *Fordham Urb. L. J.*, *28*, 457.

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.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica*, *90*(3), 1063–1084.

Fenizia, A., & Kirchmaier, T. (2025). *Not incentivized yet efficient: Working from home in the public sector*.

Finkelstein, A., Gentzkow, M., & Williams, H. (2021). Place-based drivers of mortality: Evidence from migration. *American Economic Review*, *111*(8), 2697–2735.

Frandsen, B., Lefgren, L., & Leslie, E. (2023). Judging judge fixed effects. *American Economic Review*, *113*(1), 253–277.

Frandsen, B., Lefgren, L., Leslie, E., & McIntyre, S. (2025). *The incredible shrinking instruments: Using empirical bayes to increase efficiency in judge IV designs*.

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*.

Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the new york city police department's stop-and-frisk policy in the context of claims of racial bias. *Journal of the American Statistical Association*, *102*(479), 813–823.

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

Goncalves, F., & Mello, S. (2023). *Police discretion and public safety* (31678). National Bureau of Economic Research. https://www.nber.org/papers/w31678

Goncalves, F., Mello, S., & Weisburst, E. K. (2025). *Selection bias and racial disparities in police use of force* (34175). National Bureau of Economic Research. https://www.nber.org/papers/w34175

Gonzalez, R., & Komisarow, S. (2020). Community monitoring and crime: Evidence from chicago's safe passage program. *Journal of Public Economics*, *191*, 104250.

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

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.

Hoffman, M., & Stanton, C. T. (2024). *People, practices, and productivity: A review of new advances in personnel economics* (32849). National Bureau of Economic Research. https://www.nber.org/papers/w32849

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

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., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*(1), 801–825.

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

Keane, M., & Neal, T. (2022). *A practical guide to weak instruments* (Discussion Paper 2021-05d). School of Economics, The University of New South Wales.

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.

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

LASD. (2019). *LASD MENTAL EVALUATION TEAMS ANNUAL REPORT*. Los Angeles

County Sheriff's Department.

Lee, D. S., & McCrary, J. (2016). *The deterrence effect of prison: Dynamic theory and evidence* (p. 59).

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.

Lochner, L. (2007). Individual perceptions of the criminal justice system. *American Economic Review*, *97*(1), 444–460.

Mastrobuoni, G. (2019). Police disruption and performance: Evidence from recurrent redeployments within a city. *Journal of Public Economics*, *176*, 18–31.

McCrary, J. (2007). The effect of court-ordered hiring quotas on the composition and quality of police. *The American Economic Review*, *97*(1), 318–353.

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

Mello, S. (2024). Fines and financial wellbeing. *The Review of Economic Studies*.

Miller, A. R., & Segal, C. (2018). Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence. *The Review of Economic Studies*, *86*(5), 2220–2247.

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.

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

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

Norris, S., Pecenco, M., & Weaver, J. (2024). The effect of incarceration on mortality. *The Review of Economics and Statistics*, *106*(4), 956–973.

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., & Ba, B. (2021). The economics of policing and public safety. *Journal of Economic Perspectives*, *35*(4), 3–28.

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

Petek, N., & Pope, N. G. (2023). The multidimensional impact of teachers on students. *Journal of Political Economy*, *131*(4), 1057–1107.

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

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

Prendergast, C. (2021). *'Drive and wave': The response to LAPD police reforms after rampart* [Working Paper].

Reeves, J. (2024). *Multitasking, incentives, and police officer behavior*.

Rivera, R. (2023). *Release, detain, or surveil?* [Working Paper]. Working Paper.

Rivera, R. (2025). Do peers matter in the police academy? *American Economic Journal: Applied Economics*, *17*(2), 127–164.

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.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*(1), 43–74.

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

Rose, E. K., & Shem-Tov, Y. (2021). How does incarceration affect reoffending? Estimating the dose-response function. *Journal of Political Economy*.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student

achievement. *The Quarterly Journal of Economics*, *125*(1), 175–214.

Rouhani, S., Tomko, C., Silberzahn, B., Sherman, S. G., & Cepeda, J. (2023). *Estimating economic costs of prosecuting simple drug possession in baltimore city*.

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

Rozema, K., & Schanzenbach, M. (2023). Does discipline decrease police misconduct? Evidence from chicago civilian allegations. *American Economic Journal: Applied Economics*.

Sanderson, E., & Windmeijer, F. (2016). A weak instrument f-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212–221.

Silver, D. (2021). Haste or waste? Peer pressure and productivity in the emergency department. *The Review of Economic Studies*, *88*(3), 1385–1417.

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

Weidmann, B., Vecci, J., Said, F., Deming, D. J., & Bhalotra, S. R. (2024). *How do you find a good manager?* (32699). National Bureau of Economic Research.

Weisburd, S. (2021). Police presence, rapid response rates, and crime prevention. *The Review of Economics and Statistics*, *103*(2), 280–293.

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*.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445.

Xu, G. (2018). The costs of patronage: Evidence from the british empire. *American Economic Review*, *108*(11), 3170–3198.

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	0.477 (0.0975)	0.615 (0.147)	0.594 (0.148)	0.597 (0.145)
Assignment				
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: Activities (per 900 Officer-Hours)				
Street Stop	34.7 (111)	35.9 (111)	57.2 (148)	14.3 (59.6)
Traffic Stop	12.3 (55.5)	7.38 (41.9)	5.13 (37)	22.9 (74.4)
Arrest NonIndex	10.3 (37.2)	12.9 (43.3)	10.8 (38.2)	7.53 (29.4)
Arrest Index	1.51 (13.6)	1.61 (14.1)	1.58 (14.1)	1.36 (12.8)
Force	0.348 (6.04)	0.398 (6.49)	0.365 (6.2)	0.288 (5.47)
Complaints	0.169 (4.14)	0.25 (5.03)	0.201 (4.52)	0.0705 (2.67)
Activity Cost ($1,000)	76.9 (290)	93.3 (322)	80.9 (299)	59 (248)

Note: Table presents summary statistics for officer-shift level observations, with means and standard deviations (in parentheses). Geographic units refer to those which map directly to police districts, and Movers refers to officers which switch between assignments at least once. Panel A computes officer-level statistics then takes the observation-weighted average over officers, Panel B computes counts over all officer-shift observations, and Panel C averages activities per hour across officer-shift observations.

Table 2: Summary Statistics of Teams Data

	Pooled	2010-2012	2013-2015	2016-2019
	(1)	(2)	(3)	(4)
A: By Officer				
N. Teams (Obs.)	7897 (5882)	3783 (1604)	3890 (1815)	3858 (2609)
N. Unique Units	1.57 (0.904)	1.3 (0.609)	1.28 (0.601)	1.26 (0.57)
B: Pooled				
Total Teams (District-Hour Obs.)	1835819	584709	539480	711630
Unique POs	11468	7186	7338	9029
Unique Units	25	25	22	22
Team Size (N. Active POs)	49.3 (18.2)	46.5 (17.7)	52.9 (19.1)	48.9 (17.4)
C: Crimes Reported				
Index Violent	0.132 (0.381)	0.139 (0.393)	0.12 (0.362)	0.136 (0.384)
Murder	0.0026 (0.0532)	0.00214 (0.0485)	0.0023 (0.0494)	0.00321 (0.0594)
Crim. SA	0.00826 (0.0963)	0.00746 (0.0904)	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	0.338 (0.651)	0.355 (0.656)	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 (0.584)	0.321 (0.612)	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 (0.709)	0.437 (0.729)	0.415 (0.707)	0.407 (0.695)
D: Policing Activities				
N. Street Stops	1.36 (2.46)	1.43 (2.26)	2.43 (3.37)	0.488 (1.11)
N. Traffic Stops	1.18 (2.31)	0.645 (1.14)	0.526 (1.08)	2.12 (3.22)
N. Nonindex Arrests	0.44 (0.89)	0.531 (1)	0.481 (0.944)	0.333 (0.721)
N. Index Arrests	0.0676 (0.293)	0.0685 (0.296)	0.0727 (0.306)	0.0629 (0.281)
N. Uses of Force	0.0191 (0.143)	0.0217 (0.154)	0.0206 (0.148)	0.0158 (0.13)
N. Complaints	0.0129 (0.204)	0.0183 (0.26)	0.0197 (0.245)	0.00322 (0.0793)
E: Costs ($1,000)				
Activity Cost	3.48 (7.44)	4.08 (8.28)	3.86 (7.99)	2.71 (6.09)
Crime Cost (Low Murder)	22.1 (51.8)	22.9 (49.9)	20.8 (51)	22.5 (53.9)
Crime Cost (Full Murder)	41.5 (420)	38.8 (382)	37.9 (390)	46.5 (469)
Total Cost (Low Murder)	25.6 (53.2)	27 (51.7)	24.7 (52.5)	25.2 (54.8)
Total Cost (Full Murder)	45 (421)	42.8 (382)	41.8 (391)	49.2 (469)

Note: Table presents summary statistics of team-level data where each observation is a district-hour, with means and standard deviations (in parentheses). Panel A computes officer-level statistics then takes the observation-weighted average over officers, Panel B summarizes over the pooled data, Panels C, D, and E summarizes reported crimes, policing activities, and costs per district-hour.

Table 3: Decomposition of Shift-Level Officer Effects on Policing Activities and Cross-Activity Correlations

	Street Stops	Traffic Stops	Nonindex Arrests	Index Arrests	Use of Force	Complaints	Activity Cost ($1,000)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A: Officer and Assignment Heterogeneity							
Mean(Y) per 900 θ_i-Hours	34	13	10	1.5	0.34	0.17	76
SD(θ_i) per 900 θ_i-Hours	24	18	8.5	1.3	0.37	0.13	54
	(0.12)	(0.11)	(0.04)	(0.016)	(0.006)	(0.005)	(0.24)
Share Bias in $Var(\hat{\theta}_i)$	0.19	0.2	0.11	0.32	0.27	0.53	0.12
	(0.002)	(0.003)	(0.002)	(0.011)	(0.009)	(0.018)	(0.002)
Share Var(Y) Exp.	0.17	0.22	0.14	0.023	0.005	0.002	0.086
	(0.004)	(0.004)	(0.002)	(0.0006)	(0.0003)	(0.0004)	(0.001)
Var(θ_i) / Var(Y)	0.049	0.1	0.053	0.009	0.004	0.001	0.036
	(0.0005)	(0.001)	(0.0005)	(0.0002)	(0.0001)	(0.00008)	(0.0003)
Var($\psi_{J,P}$) / Var(Y)	0.12	0.083	0.069	0.014	0.002	0.001	0.038
	(0.004)	(0.002)	(0.002)	(0.0006)	(0.0003)	(0.0004)	(0.002)
2 x Cov(θ_i, $\psi_{J,P}$) / Var(Y)	0.009	0.033	0.015	0.0005	-0.0003	0.0000009	0.013
	(0.0009)	(0.002)	(0.0007)	(0.0005)	(0.0002)	(0.00007)	(0.0006)
Cor(θ_i, $\psi_{J,P}$)	0.061	0.18	0.13	0.024	-0.052	0.0004	0.18
	(0.006)	(0.01)	(0.008)	(0.021)	(0.029)	(0.027)	(0.011)
B: Cross-Activity Correlation between Officer Effects							
Street Stops	-	0.285	0.521	0.428	0.231	0.0668	0.532
	-	(0.0034)	(0.0043)	(0.0076)	(0.017)	(0.032)	(0.0041)
Traffic Stops	-	-	0.25	0.192	0.131	0.113	0.254
	-	-	(0.004)	(0.0066)	(0.013)	(0.019)	(0.0043)
Nonindex Arrests	-	-	-	0.489	0.38	0.268	0.96
	-	-	-	(0.0096)	(0.014)	(0.021)	(0.0015)
Index Arrests	-	-	-	-	0.333	0.121	0.67
	-	-	-	-	(0.013)	(0.02)	(0.0058)
Uses of Force	-	-	-	-	-	0.632	0.54
	-	-	-	-	-	(0.028)	(0.0089)
Complaints	-	-	-	-	-	-	0.382
	-	-	-	-	-	-	(0.018)

Note: Table summarizes the variance decomposition of estimated effects in equation (1) in Panel A and the cross-outcome correlation between officer effects estimated in equation (1) in Panel B. Standard errors computed using nonparametric bootstrap (40 runs) are below estimates in parentheses. All variance and covariance components are bias-corrected using KSS, and similarly for the cross-activity correlations following Lachowska et al. (2023).

Table 4: Decomposition of Team-Level Officer Effects on Crime and Activities Results

	Mean(Y) x900	$\widehat{SD(\omega_i)}$ x900	$\widehat{Var(\omega_i)}/Var(Y)$ x 100	Share Plug-In Bias
	(1)	(2)	(3)	(4)
A: Crime				
Index Violent	119	2.13	0.004	0.942
		(0.373)	(0.001)	(0.0193)
Index Property	442	6.12	0.007	0.869
		(0.426)	(0.001)	(0.018)
Nonindex Drug	111	4.41	0.0119	0.782
		(0.161)	(0.0009)	(0.0158)
Nonindex Simple Violent	272	4.25	0.006	0.901
		(0.343)	(0.0009)	(0.0157)
Nonindex Weapon	18.3	0.461	0.0004	0.995
		(0.218)	(0.001)	(0.0221)
Nonindex All Other	377	5.17	0.006	0.896
		(0.416)	(0.001)	(0.0165)
B: Police Activity				
N. Street Stops	1220	40.6	0.0268	0.515
		(0.403)	(0.0005)	(0.01)
N. Traffic Stops	1060	54	0.0557	0.369
		(0.449)	(0.0009)	(0.0105)
N. Nonindex Arrests	396	8.92	0.0106	0.801
		(0.391)	(0.0009)	(0.0173)
N. Index Arrests	60.8	1.86	0.004	0.923
		(0.25)	(0.001)	(0.0191)
N. Uses of Force	17.2	0.99	0.005	0.919
		(0.103)	(0.001)	(0.0165)
N. Complaints	11.6	0.942	0.002	0.959
		(0.228)	(0.001)	(0.0222)
C: Costs				
Activity Cost	3140	70.8	0.01	0.83
		(3.77)	(0.001)	(0.0175)
Crime Cost (Low Murder)	19900	330	0.004	0.929
		(42.1)	(0.001)	(0.0175)
Total Cost (Low Murder)	23000	358	0.005	0.92
		(40.6)	(0.001)	(0.0175)

Note: Table summarizes the variance decomposition of estimated officer effects in equation (2). All variance components are bias-corrected using the split-sample procedure averaged over 40 draws, and standard errors in parentheses are computed as the standard deviation of the components across the draws.

Table 5: Effect of Policing Activities on Crimes per District-Hour

	Reduced Form (Officer Activity Effects)		First Stage		2SLS		Elasticities (2SLS)		OLS	
	Violent	Property	Focal Instrument	Sanderson-Windmeijer F-Stat	Violent	Property	Violent	Property	Violent	Property
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Street Stops	-0.001	-0.01***	0.39***	340.5	-0.003	-0.03***	-0.028	-0.082***	-0.0002	0.0006
	(0.001)	(0.003)	(0.01)		(0.004)	(0.007)	(0.041)	(0.021)	(0.0002)	(0.0004)
Traffic Stops	0.0009	-0.009*	0.81***	1084.8	0.002	-0.01	0.015	-0.023	-0.004***	-0.002***
	(0.003)	(0.005)	(0.02)		(0.003)	(0.006)	(0.029)	(0.015)	(0.0002)	(0.0004)
Nonindex Arrests	-0.009	-0.02*	0.34***	189.9	-0.026	-0.067**	-0.086	-0.06**	0.005***	0.005***
	(0.006)	(0.011)	(0.02)		(0.019)	(0.033)	(0.063)	(0.03)	(0.0005)	(0.0009)
Index Arrests	0.012	0.033	0.2***	45.5	-0.00009	0.059	-0.00005	0.008	0.053***	0.2***
	(0.029)	(0.063)	(0.03)		(0.15)	(0.32)	(0.076)	(0.044)	(0.001)	(0.003)
Uses of Force	-0.044	-0.072	0.57***	91	0.008	-0.26	0.001	-0.01	0.077***	0.002
	(0.12)	(0.2)	(0.05)		(0.23)	(0.41)	(0.034)	(0.016)	(0.003)	(0.005)
Complaints	0.16	-0.47	0.86***	32.7	0.2	-0.54	0.02	-0.014	0.004**	0.004
	(0.22)	(0.38)	(0.15)		(0.26)	(0.47)	(0.026)	(0.012)	(0.002)	(0.003)

Note: Table displays estimates of the effect of officer propensities for policing activities and marginal policing activities on violent and property crime at the district-hour level using equations (5)-(7). Columns (1)-(2) display reduced form relationships officer activity effects and crime, interpreted as the effect of a 1 unit increase in team propensity for the specific activity (e.g., increasing a single team members' propensity for street stops by 1 per hour), estimated from equation (5). Column (3) displays the first stage relationship between the number of policing activities at the district-hour level and the sum of officer fixed effects for that activity working on the team in that district-hour (equation (7)). Column (4) displays the Sanderson-Windmeijer F-statistics which account for multiple endogenous variables. Column (5)-(8) are 2SLS estimates (equations (6) and (7)), with Columns (5) and (6) being the effect of the marginal activity, and Columns (7) and (8) being the implied elasticites of crime with respect to that activity, which multiplies the 2SLS coefficient by the mean of the policing activity per district-hour and divides it by the mean of the crime type per district-hour. Columns (9) and (10) are the OLS estimates using only equation (6). Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effects of Policing Activities on Costs ($1,000)

	Reduced Form (λ_k)			2SLS (η_k)			Implied Regression (β_k)		
	Activity	Crime	Total	Activity	Crime	Total	Activity	Crime	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Street Stops	-0.077**	-0.91***	-0.98***	<0.001	-2.35***	-2.35***	-0.03	-1.54	-1.9
	(0.03)	(0.22)	(0.22)	(<0.001)	(0.581)	(0.581)	(-)	(1.78)	(1.66)
Traffic Stops	0.062	0.45	0.51	<0.001	0.679	0.679	-0.05	-0.18	-0.33
	(0.043)	(0.36)	(0.37)	(<0.001)	(0.459)	(0.459)	(-)	(0.98)	(1.03)
Nonindex Arrests	1.7***	-0.94	0.72	5***	-3.4	1.6	5.38	5.62	15.08
	(0.13)	(0.84)	(0.86)	(<0.001)	(2.55)	(2.55)	(-)	(15.78)	(17.71)
Index Arrests	-0.13	1.1	0.93	10***	-0.566	9.43	12.62	17.77	15.09
	(0.63)	(4)	(4.1)	(<0.001)	(20.4)	(20.4)	(-)	(86.86)	(97.9)
Uses of Force	18***	7.3	26	23.9***	17.2	41.1	14.13	77.02	76.13
	(2.4)	(17)	(17)	(<0.001)	(31.8)	(31.8)	(-)	(190.91)	(198.76)
Complaints	15***	1.9	17	12***	2.68	14.7	-2.54	-22.27	0.88
	(4.3)	(30)	(31)	(<0.001)	(36.6)	(36.6)	(-)	(180.92)	(204.78)

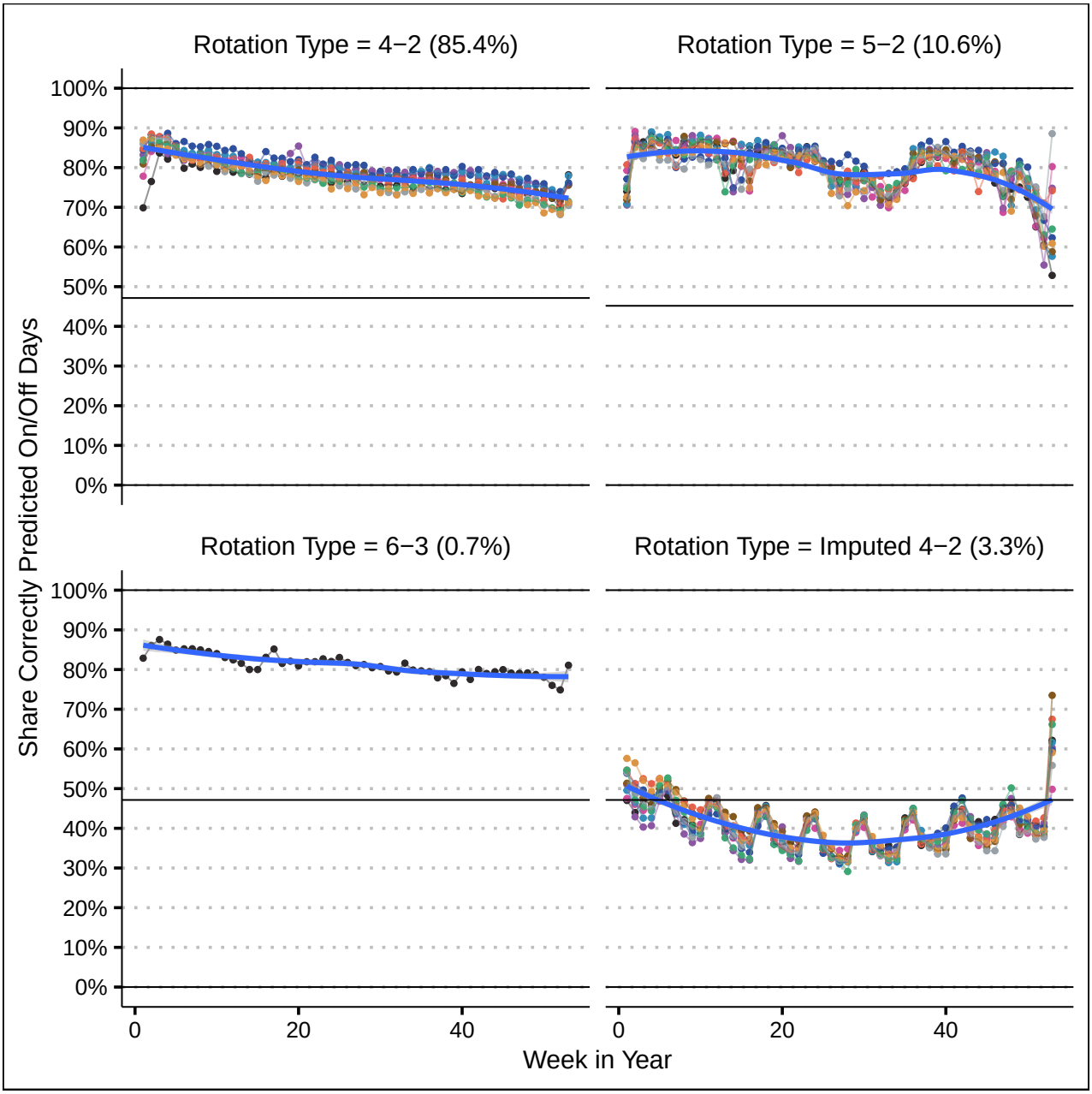
Note: Table displays coefficients of the reduced form relationship between activity, crime, and total costs at the district-hour level and officer heterogeneity for policing activities using different designs. Columns (1)-(3) are the reduced form effects of team propensities on costs (equation (5)); Columns (4)-(6) are the 2SLS estimates for the effect of the marginal activity on costs (equations (6)-(7)); and Columns (7)-(9) are the implied regression results of projecting team-level officer effects for costs onto team-level officer effects for activities (equation (4)). Crime costs refer to total crime costs using Miller et al. (2021) incidence costs applied to reported crimes but using the 'Low Murder' cost (valuing a murder at $400,000 rather than $8 million). Standard errors are in parentheses, and in Columns (1)-(6) are clustered at the district-year-month-day-of-week-hour level while in Columns (7)-(9) are computed as standard deviations across split-sample draws. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Projecting Officer Effects onto Officer Demographics (per 900 Officer-Hours)

	Activity Effects ($\hat{\theta}$)						Crime and Cost Effects ($\hat{\omega}$)			
	Street Stops	Traffic Stops	Nonindex Arrests	Index Arrests	Use of Force	Complaints	Violent Crimes	Property Crimes	Crime Cost ($1,000)	Total Cost ($1,000)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	-11***	-1.7***	-2.4***	-0.22***	-0.12***	-0.028***	0.029	-0.19	-18	-22
	(0.46)	(0.43)	(0.12)	(0.035)	(0.017)	(0.0075)	(0.23)	(0.43)	(31)	(32)
Hispanic	-0.97***	4.9***	0.18*	0.096**	-0.051***	-0.012**	0.09	-0.026	0.17	-5
	(0.37)	(0.38)	(0.1)	(0.04)	(0.015)	(0.0058)	(0.24)	(0.48)	(35)	(35)
Other Race	-2.6***	3.6***	-0.68***	0.05	-0.095***	-0.025***	0.54	-0.36	100*	89
	(0.65)	(0.73)	(0.21)	(0.058)	(0.018)	(0.0084)	(0.39)	(0.79)	(60)	(61)
Female	-4.5***	0.72**	-2.2***	-0.17***	-0.16***	-0.048***	-0.045	1.2***	-18	-18
	(0.29)	(0.3)	(0.077)	(0.034)	(0.021)	(0.0048)	(0.21)	(0.42)	(28)	(29)
S. Age = [25,28]	-0.84**	-2.2***	-0.36***	-0.13***	-0.063**	-0.0012	0.11	0.14	-9.5	-7.4
	(0.34)	(0.36)	(0.11)	(0.038)	(0.027)	(0.0055)	(0.25)	(0.5)	(35)	(36)
S. Age = [29,32]	-2.1***	-2.6***	-0.85***	-0.19***	-0.083***	-0.0099	0.25	-0.27	-26	-23
	(0.38)	(0.4)	(0.12)	(0.04)	(0.028)	(0.0062)	(0.27)	(0.54)	(39)	(40)
S. Age = >33	-3.3***	-3.2***	-2.1***	-0.31***	-0.16***	-0.033***	0.044	0.24	-48	-56
	(0.39)	(0.41)	(0.12)	(0.04)	(0.027)	(0.0063)	(0.27)	(0.54)	(39)	(39)
Spanish Speaking	-2.7***	-4.7***	-1.1***	-0.17***	-0.015	0.017**	-0.072	-0.69	-27	-25
	(0.4)	(0.38)	(0.11)	(0.039)	(0.011)	(0.0076)	(0.28)	(0.57)	(38)	(38)
Edu >= BA	4.9***	5.8***	0.77***	0.13***	0.027*	-0.0088*	0.26	0.55	14	17
	(0.31)	(0.28)	(0.084)	(0.026)	(0.016)	(0.0051)	(0.21)	(0.4)	(30)	(31)
Missing Edu	-0.42	5.4***	0.15	-0.094**	-0.01	0.0012	0.24	0.32	27	29
	(0.3)	(0.35)	(0.1)	(0.038)	(0.013)	(0.0057)	(0.22)	(0.46)	(32)	(33)
Tenure 5-10	-5***	-17***	-0.46***	-0.16***	-0.022	0.013*	0.26	0.077	36	49
	(0.5)	(0.32)	(0.14)	(0.044)	(0.017)	(0.0075)	(0.27)	(0.47)	(40)	(41)
Tenure 10-15	-12***	-19***	-3.2***	-0.55***	-0.1***	-0.0072	0.46	0.97*	46	55
	(0.53)	(0.32)	(0.15)	(0.038)	(0.017)	(0.0075)	(0.29)	(0.58)	(39)	(40)
Tenure 15-20	-15***	-22***	-4.8***	-0.79***	-0.18***	-0.054***	0.061	0.81	-42	-47
	(0.54)	(0.43)	(0.14)	(0.058)	(0.015)	(0.0075)	(0.29)	(0.66)	(41)	(42)
Tenure >20	-18***	-24***	-5.9***	-0.99***	-0.23***	-0.066***	-0.16	2***	-17	-22
	(0.58)	(0.6)	(0.15)	(0.061)	(0.016)	(0.0079)	(0.28)	(0.67)	(38)	(39)
Raw R2	0.14	0.13	0.12	0.077	0.039	0.026	0.0007	0.003	0.001	0.001
Corrected R2	0.17	0.16	0.14	0.11	0.054	0.055	0.012	0.023	0.014	0.015
Mean Dep.	34.2	12.62	10.15	1.51	0.34	0.17	119	442	19892	23028
N Officers	12028	12028	12028	12028	12028	12028	11463	11463	11463	11463

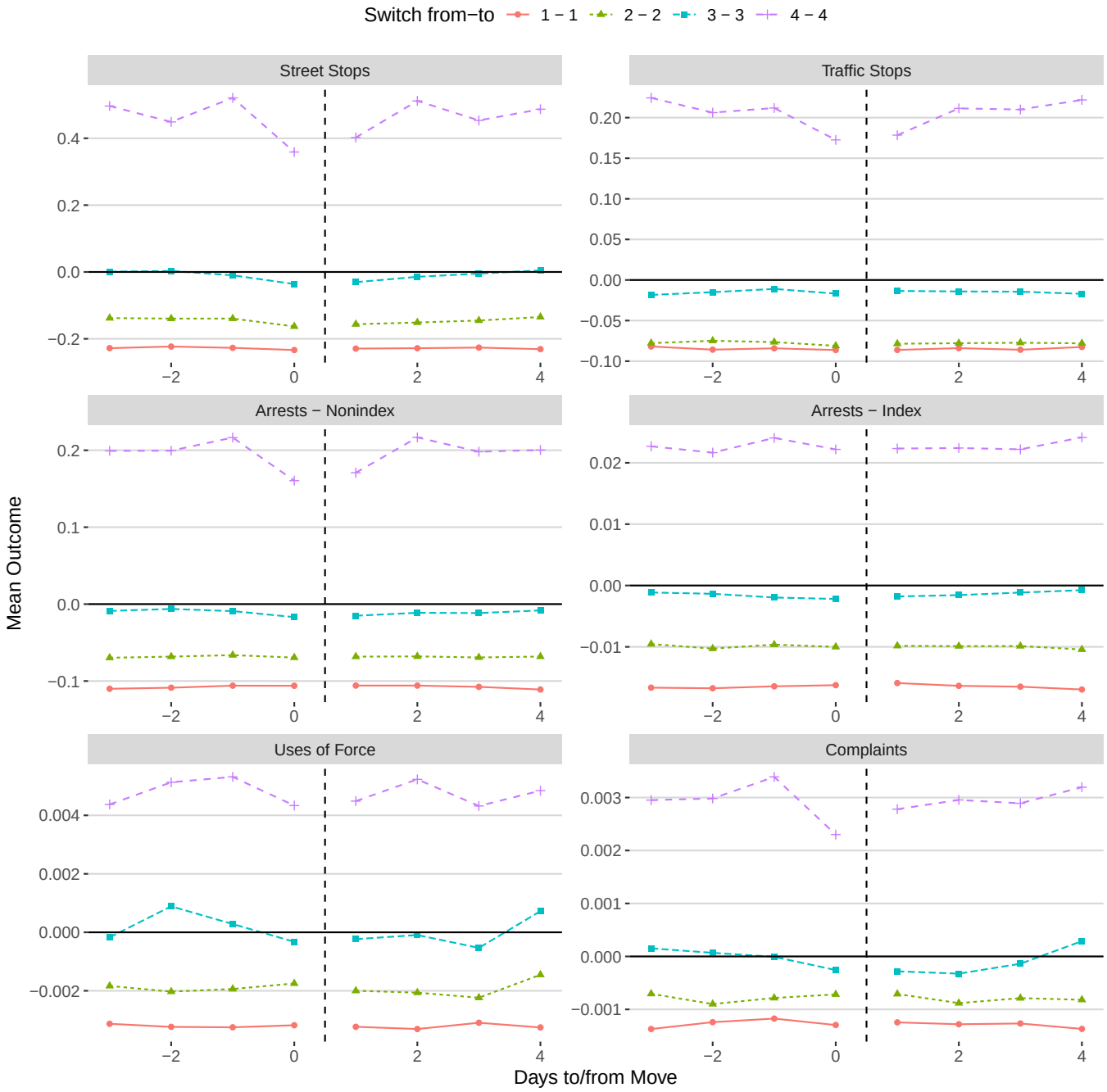
Note: Table displays the results of projecting officer effects onto officer demographics, with the reference category being officers who are white, male, started in the CPD younger than 25, do not speak Spanish, have less than a BA in education, and tenure between 1.5 and 5 years (equation (3)). Officer effects for activities are the outcomes in Column (1) - (6), recovered from estimating equation (1), and officer effects for crime and costs (low murder) are the outcomes in Columns (7)-(10), recovered from estimating equation (2). Mean Dep. refers to the mean of the initial outcome in equations (1) and (2). Standard errors in parentheses are computed: Columns (1)-(6) use KSS adjusted heteroskedasticity robust standard errors, and Columns (7)-(10) use heteroskedasticity robust standard errors (HC1). Raw R2 refers to the R-squared from the projection regression, while Corrected R2 rescales Raw R2 by the ratio of plug-in variance and bias-corrected variance in officer effects. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Accuracy of Predicted Days On and Off



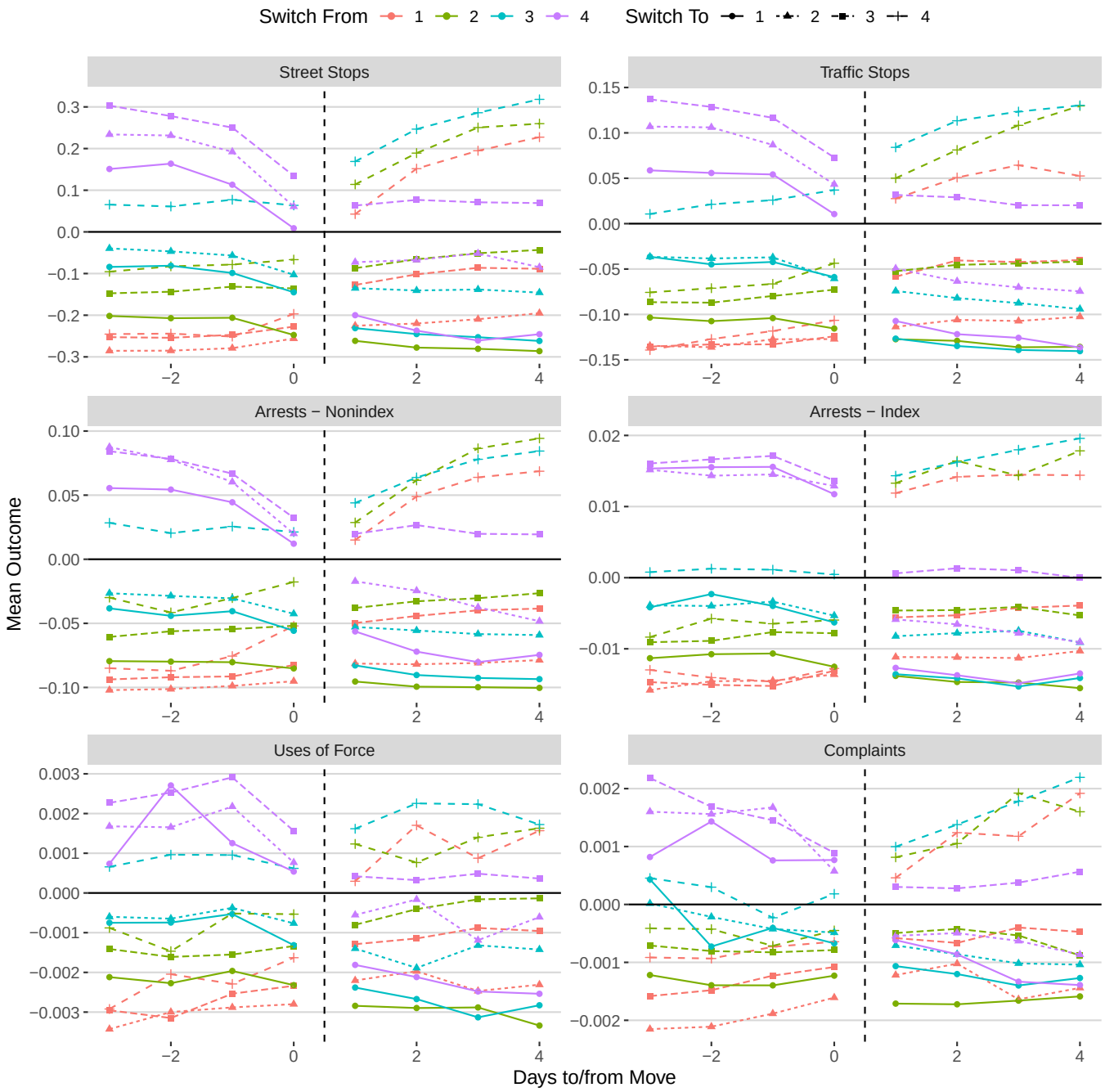
Note: Figure displays the share of correctly predicted days-on and days-off using the prediction method discussed in Section 2. The method uses officer's reported days on and off within each calendar to first identify which fixed rotation they are on (e.g., working 4 days in a row followed by 2 days off is a 4-2 rotation). Then, within each officer-year, the first time an officer follows that pattern is found (e.g., they are observed working 4 days then off for 2 days), and the pattern is projected forward and backward within that officer-year to form a predicted rotation. This predicted rotation is compared with the observed days on and off, and an officer-day is considered correctly predicted if the officer is predicted on (off) and observed to be on (off). The y-axis denotes the share of officer-day observations within a year (colors) that are correctly predicted. The x-axis denotes the calendar week within each year. Points are year-by-week correct prediction rates. The blue line denotes the nonlinear fit across years. Black horizontal line denotes the share of correctly predicted officer-days if days on (off) were randomly chosen within a spell duration (e.g., 4-2 is 6 days, 5-2 is 7 days). The Imputed 4-2 rotation type is for officer-years that did not fit any rotation type and so a 4-2 rotation is used.

Figure 2: Event Study of Policing Activities around Officers Switching Assignments



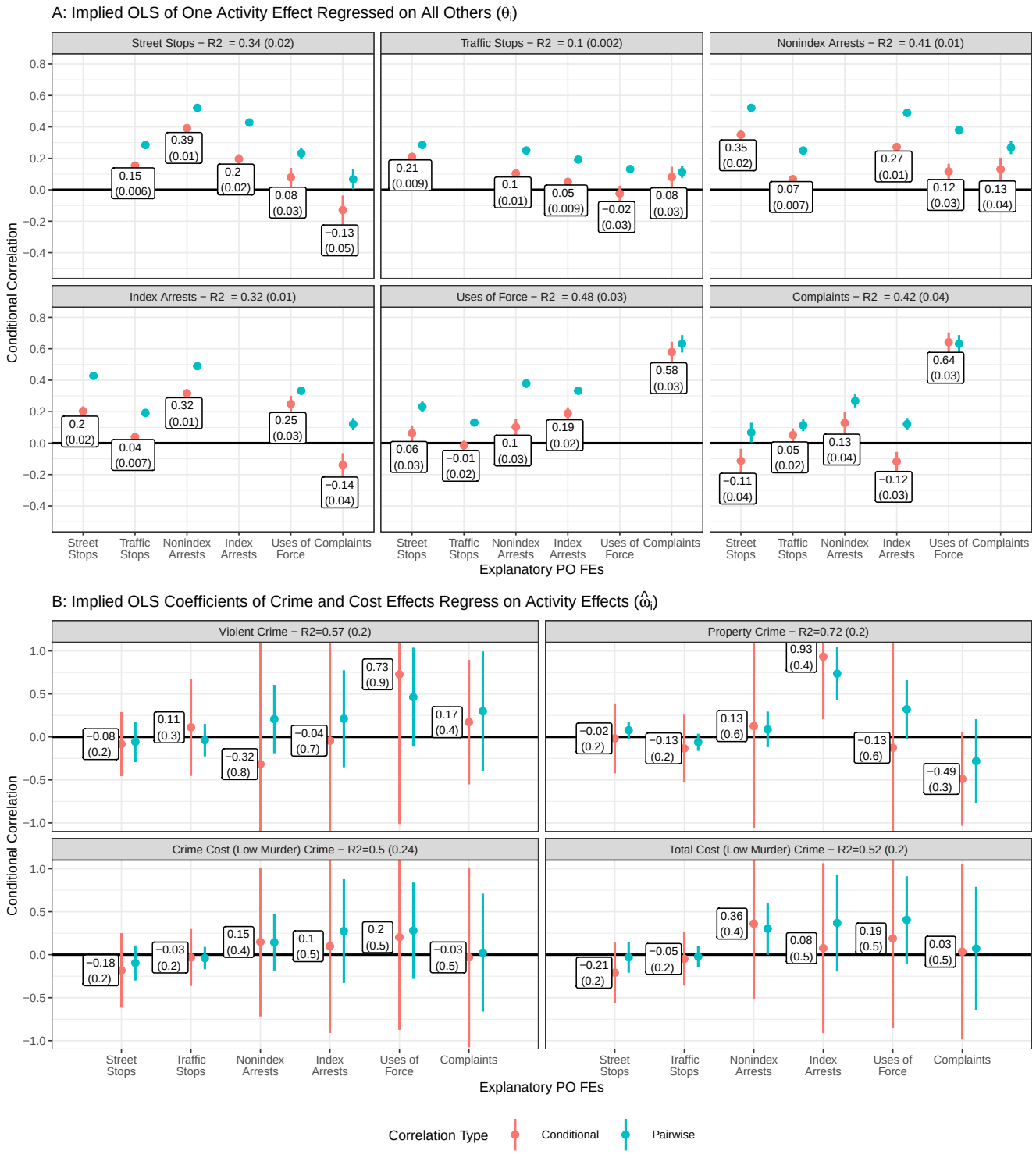
*Note:*Figure shows the time trends in policing activities around the time that officers switched between assignments. The y-axis denotes the frequency of the focal policing activity per officer-shift (de-meaned). The x-axis denotes the shifts worked before and after the switch between assignments. Points denote the average policing activity frequency for officer-assignment pairs. Officer-Assignment pairs are grouped into quantiles based on the average frequency of policing activities for the assignment within a year, excluding observations for that officer. The quantile group before the switch is denoted by the color, and the quantile group after the switch is denoted by the line type and point shapes (4 being the highest quantile, 1 being the lowest quantile). Figure only displays transitions for officers switching to/from assignments in the same quantile. 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, because officer-assignment spells vary and can be as short as 1 shift.

Figure 3: Event Study of Policing Activities around Officers Switching Assignments (Cont.)



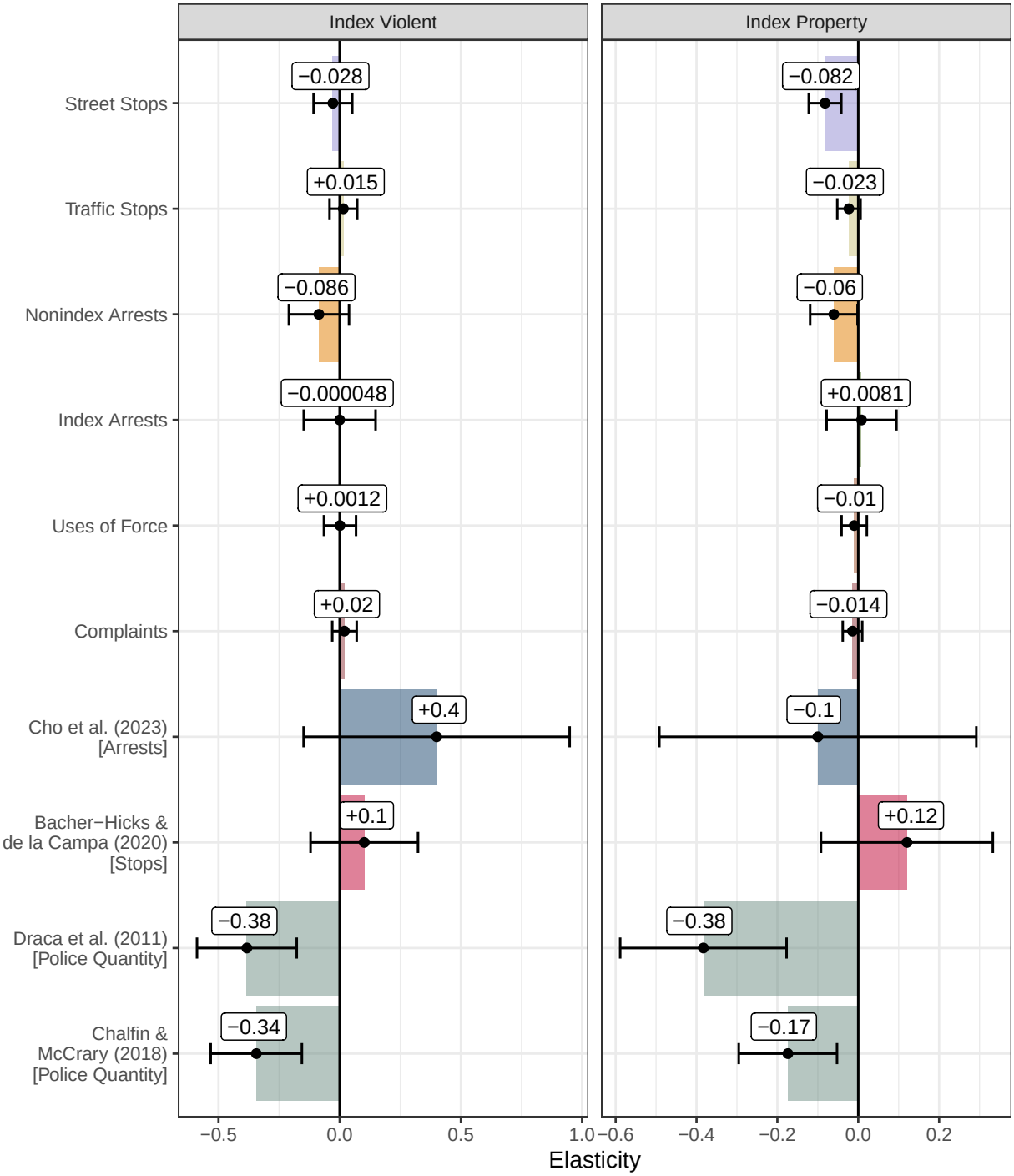
Note: Figure shows the time trends in policing activities around the time that officers switched between assignments. The y-axis denotes the frequency of the focal policing activity per officer-shift (de-meaned). The x-axis denotes the shifts worked before and after the switch between assignments. Points denote the average policing activity frequency for officer-assignment pairs. Officer-Assignment pairs are grouped into quantiles based on the average frequency of policing activities for the assignment within a year, excluding observations for that officer. The quantile group before the switch is denoted by the color, and the quantile group after the switch is denoted by the line type and point shapes (4 being the highest quantile, 1 being the lowest quantile). Figure only displays transitions for officers switching to/from assignments in different quantiles. 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, because officer-assignment spells vary and can be as short as 1 shift.

Figure 4: Effects of Officer Activity Effects and Officer Activity, Crime, and Cost Effects



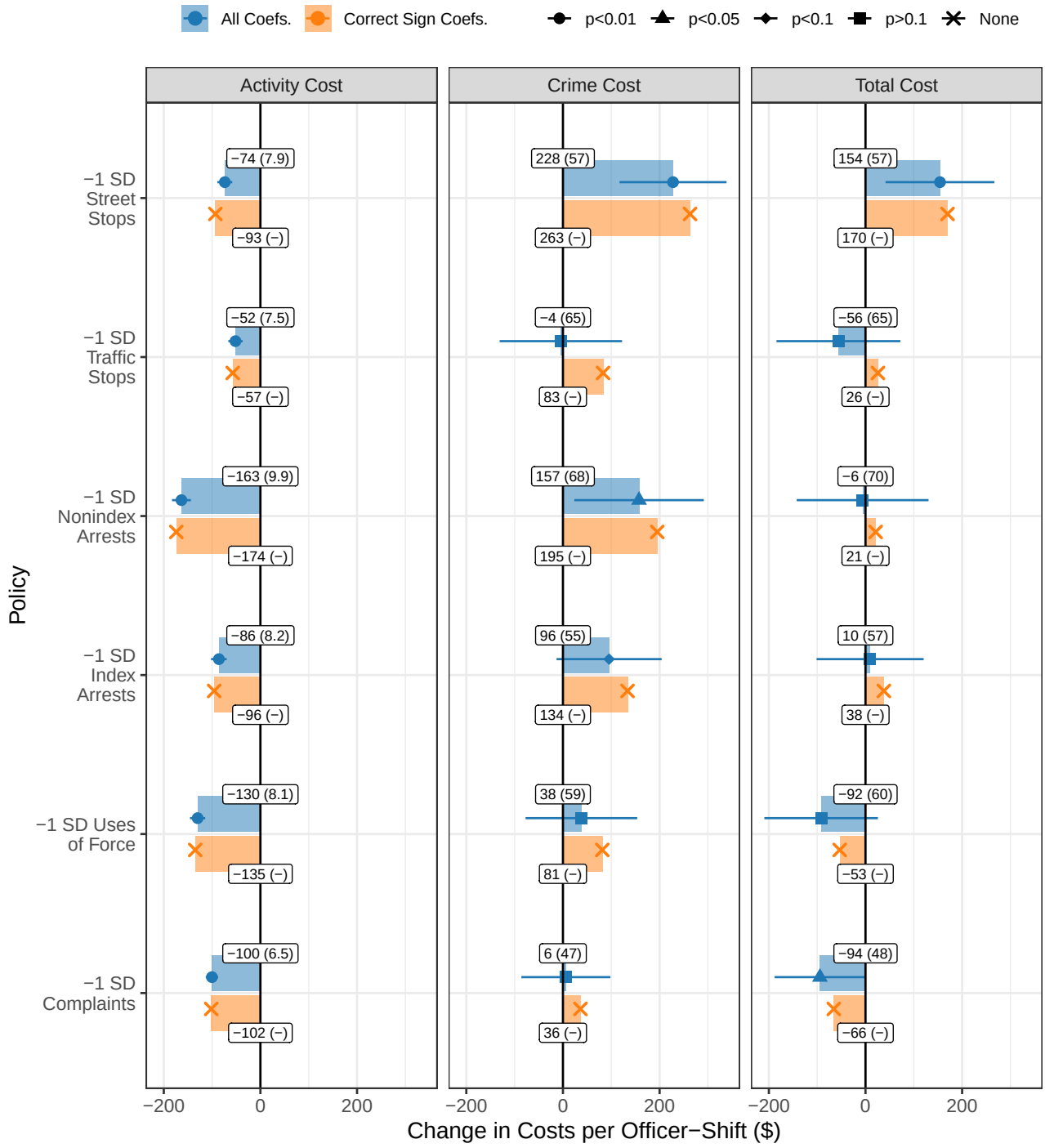
Note: Figure displays the implied regression results equivalent to regressing officer effect for one outcome on officer effects for other outcomes (explanatory variables), and compares these coefficients with pairwise correlations between the dependent variable officer effect and one of the explanatory officer effects. Implied regression estimates (equation (4)) are computed using the bias-corrected (co)variance structures and converted into 'Conditional Correlations' and are displayed next to bias-corrected pairwise correlations. Panel A uses shift-level officer effects with KSS-bias corrected (co)variances and re-scales estimates to conditional correlations as $\hat{\beta}^l_k \times \widehat{SD(\theta^k_i)}/\widehat{SD(\theta^l_i)}$). Panel B uses team-level officer effects with split-sample bias corrected (co)variances and re-scales estimates to conditional correlations as $\hat{\beta}^c_k \times \widehat{SD(\omega^k_i)}/\widehat{SD(\omega^c_i)}$). Labels display resulting R^2 of the implied regressions. Bootstrapped standard errors are displayed in parentheses and used to compute the 95% confidence intervals.

Figure 5: Elasticities of Crime with Respect to Policing Activities and Quantity



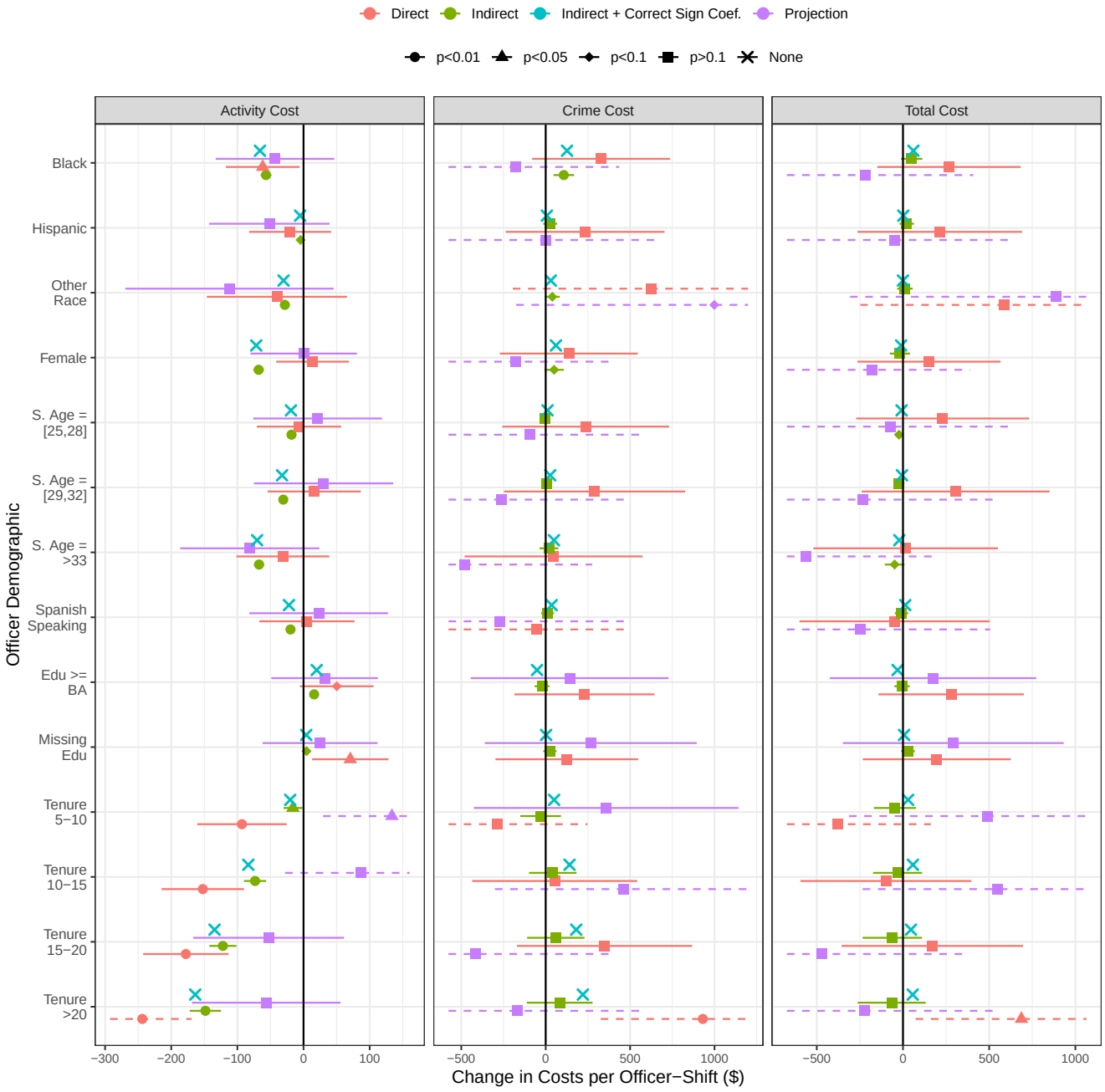
Note: Figure compares elasticities of violent and property crime with respect to policing activities estimated in this paper with elasticities with respect to arrests, stops, and police quantity from prior work. Elasticities in this paper (from Columns (7) and (8) of Table 5) are computed from the 2SLS coefficients of the effect of each policing activity on crime at the district-hour level (equations (6) and (7)), then multiplying coefficients by the mean of the policing activity per district-hour and dividing it by the mean of the crime type per district-hour. 95% confidence intervals are based on standard errors clustered at the district-year-month-day-of-week-hour level. External sources: Cho et al. (2023) Figure 8; Bacher-Hicks and de la Campa (2020b) uses effects of 1 SD in commander stop effects on violent and property felony estimates (Table A8) divided their respective mean divided by 1 SD in commander stop effects divided by the mean number of stops, note the standard errors do not take into account first stage imprecision; Draca et al. (2011) uses the IV estimates for susceptible crimes (including both property and violent offenses, hence they are used for both types) in Column (4) of Table 4 Part I; Chalfin and McCrary (2018) uses the violent and property crime elasticity estimates in Column (9) of Table 3.

Figure 6: Effects of Replacing Officers on Costs per Officer-Shift



Note: Figure displays the effects on costs of replacing an officer whose officer effect for a specific activity is 1 standard deviation above the mean with an average officer. Effects are scaled to be per officer 9-hour shift. Changes in costs take into account the 1 SD decline in the focal activity and the implied changes in other activities based on the bias-corrected covariances between officer effects across activities. Officer effects are computed from estimating equation (1), and standard deviations and covariances are KSS bias-corrected. Changes in activities are converted into changes in costs by applying changes in officer effects to reduced form coefficients of the effect of team propensities on costs at the district-hour level (equation (5)) with standard errors clustered at the district-year-month-day-of-week-hour level, see Columns (1)-(3) of Table 6. All Coefs. results use the reduced form effects of team propensities on activity costs, crime costs (low murder), and total costs (activity costs + crime costs), with point estimates and corresponding standard errors computed using a linear combination of the coefficients and covariance matrix. Correct Sign Coefs. removes any coefficient which is 'incorrectly signed', which means a positive effect of policing activity effects on crime costs or a negative effect of policing activity effects on activity costs (no standard errors are computed).

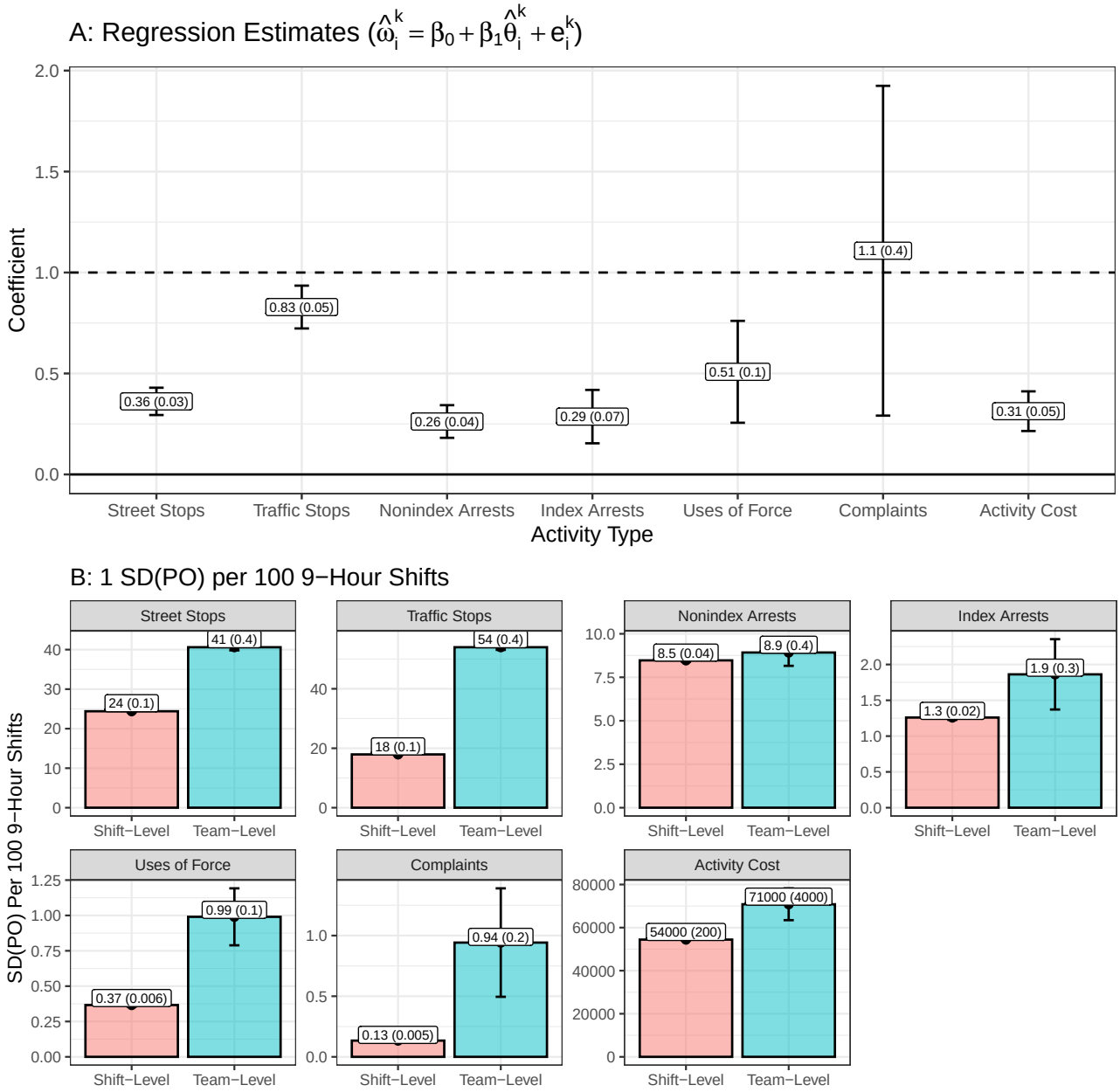
Figure 7: Effects of Officer Demographics on Costs per Officer-Shift



Note: Figure displays the effects of officer demographics on activity costs, (low murder) crime costs, and total costs (activity costs + crime costs) per 9-hour officer-shift along with 95% confidence intervals. Estimates can be interpreted as the effect of having an officer with a specific demographic on a team relative to the reference officer (white, male, started in the CPD younger than 25, do not speak Spanish, have less than a BA in education, and tenure between 1.5 and 5 years). 'Direct' regresses costs at the district-hour level onto team demographics (equation (8)), with standard errors clustered at the district-year-month-day-of-week-hour level and used to construct confidence intervals. The 'Indirect' method uses the projection coefficients from regressing shift-level officer fixed effects estimated from equation (1) on officer demographics (equation (3)) multiplied the reduce form effects of officer effects on costs (equation (5)) to construct the expected change in costs (equation (9)), with linear combinations of the variance-covariance matrix of reduce form effects used to compute standard errors. The 'Indirect + Correct Sign Coef.' method only uses the correct signed coefficients (coefficients which have activities weakly increasing activity costs and weakly decreasing crime costs) for the Indirect method. 'Projection' uses the projection coefficients recovered from regressing team-level officer effects for costs recovered from estimating equation (2) on officer demographics (equation (3)) weighted by officer-shift observations, with heteroskedasticity robust standard errors (HC1) used to compute confidence intervals.

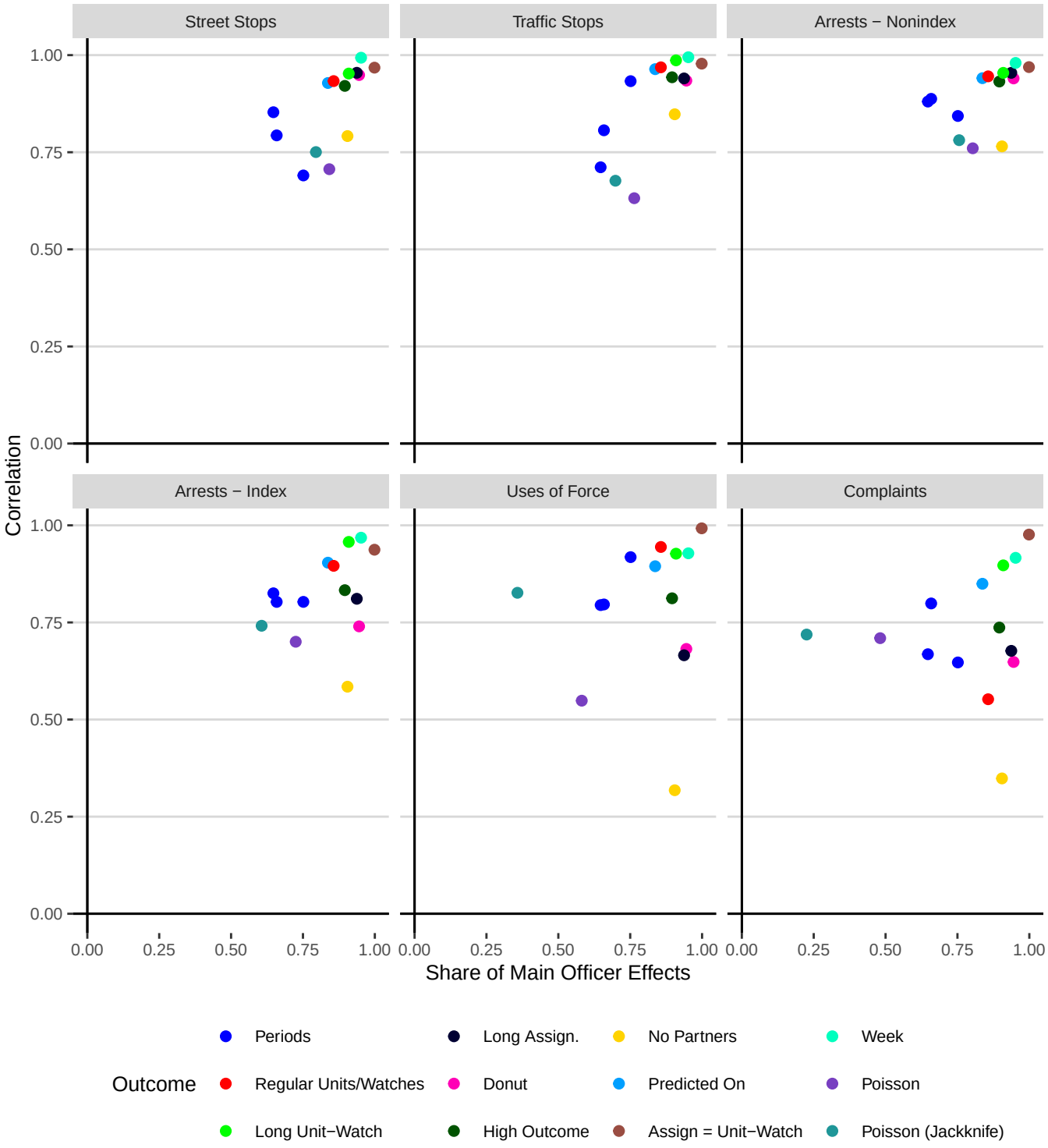
A Additional Figures and Tables

Figure A.1: Comparison of Officer Activity Effects using Shift and Team-Level Designs



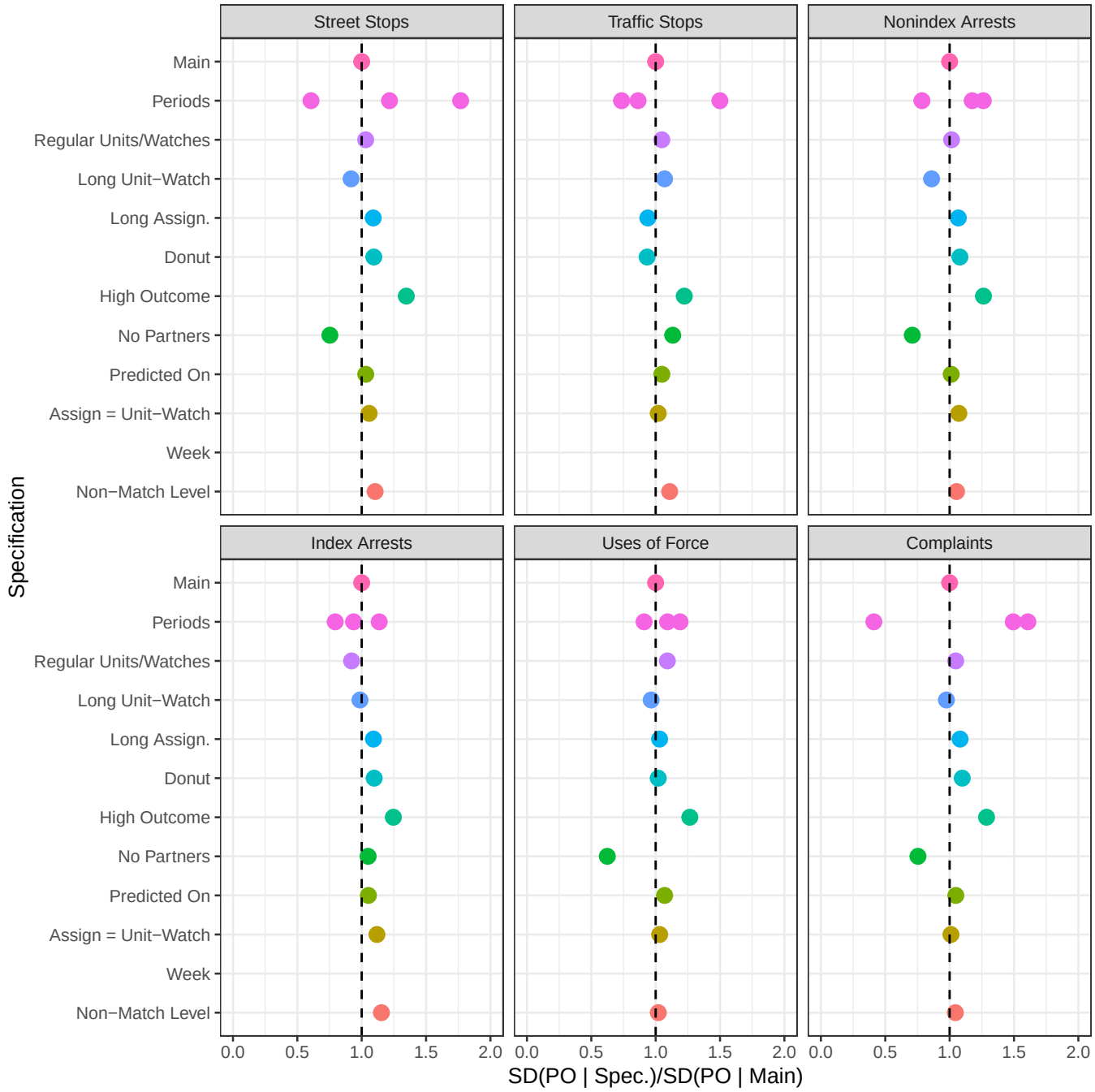
Note: Figure compares the officer activity effects estimated from the shift-level design (equation (1)) and team-level design (equation (2)). Panel A displays OLS estimates of regressing estimated team-level officer effects ($\hat{\omega}_i^k$) for an activity onto the shift-level officer effects for the same activity ($\hat{\theta}_i^k$), with regressions weighted by officer's number of shifts and standard errors clustered at the officer level used to compute 95% confidence intervals. Panel B compares the KSS bias-corrected standard deviations of officer effects from the shift-level design (bootstrap standard errors displayed and used for 95% confidence intervals) with the split-sample bias-corrected standard deviations of officer effects from the team-level design (bootstrap standard errors displayed and used for 95% confidence intervals computed from variation over parameters by draws that determine splits).

Figure A.2: Correlation between Officer Effects Across Specifications



Note: Figure displays the (non-bias corrected) correlation (y-axis) between shift-level officer effects recovered from equation (1) with officer effects recovered in other specifications and samples. The x-axis contains the share of officer effects recovered in the main specification that were also recovered in the alternative specification. See Appendix B.1 for details on the specifications.

Figure A.3: Magnitude of Officer Effect Heterogeneity by Specification



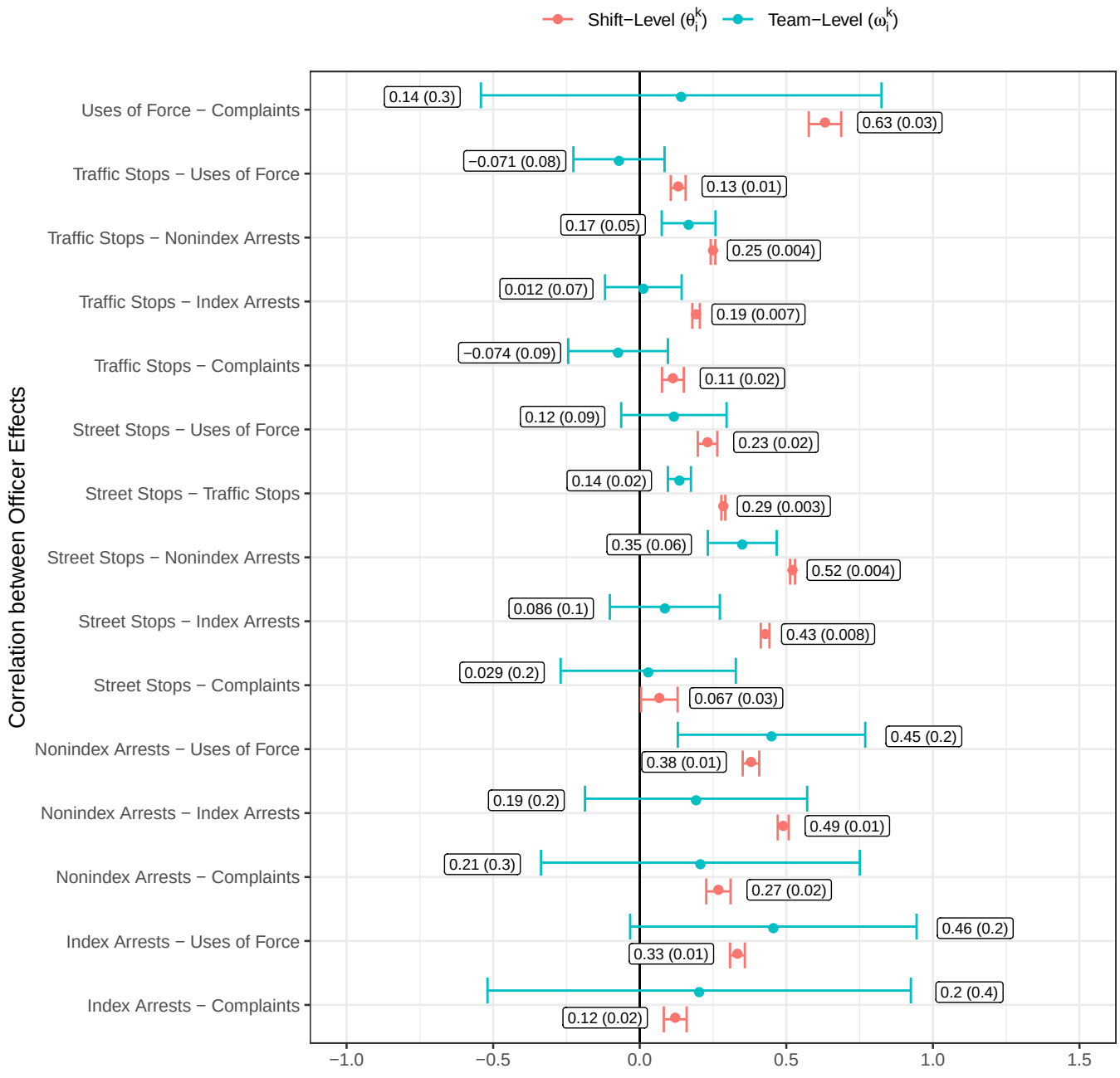
Note: Figure displays the KSS bias-corrected standard deviations of officer effects (recovered from equation (1)) relative to the main specification. See Appendix B.1 for details on the specifications.

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



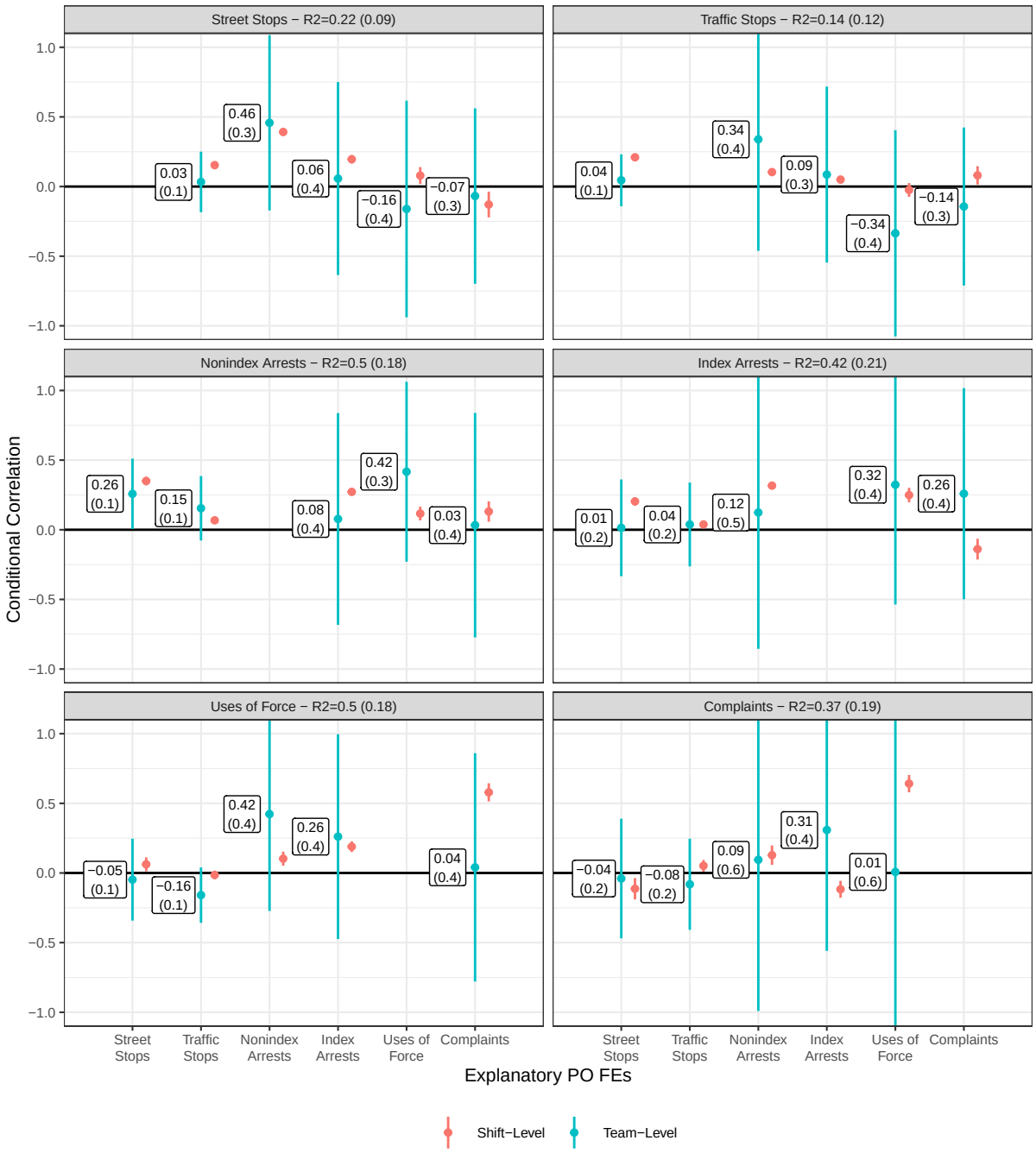
Note: Figure displays the KSS bias-corrected (excluding Poisson) cross-outcome correlations for officer effects estimated from equation (1) across different specifications. Dashed lines correspond to the value of the main specification. See Appendix B.1 for details on the specifications.

Figure A.5: Comparison of Cross-Activity Correlations in Shift-Level and Team-Level Designs



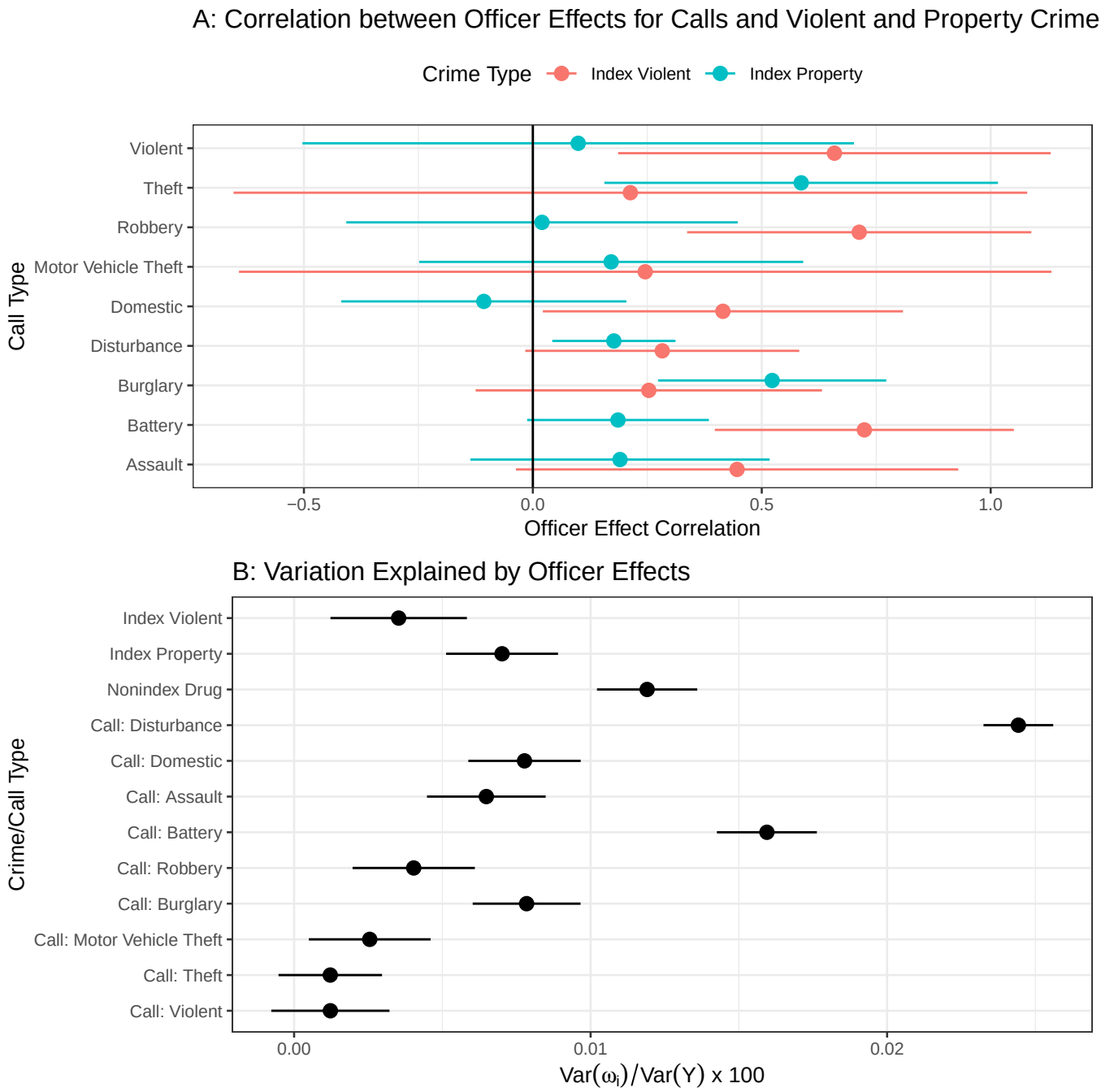
Note: Figure displays the KSS and split-sample bias-corrected cross-activity correlations in officer effects from the shift-level (equation (1)) and team-level (equation (2)) designs. Bootstrap standard errors displayed and used for 95% confidence intervals for both, with team-level bootstrap runs coinciding the draws used to compute splits.

Figure A.6: Comparison of Conditional Correlations of Officer Activity Effects in Shift-Level and Team-Level Designs



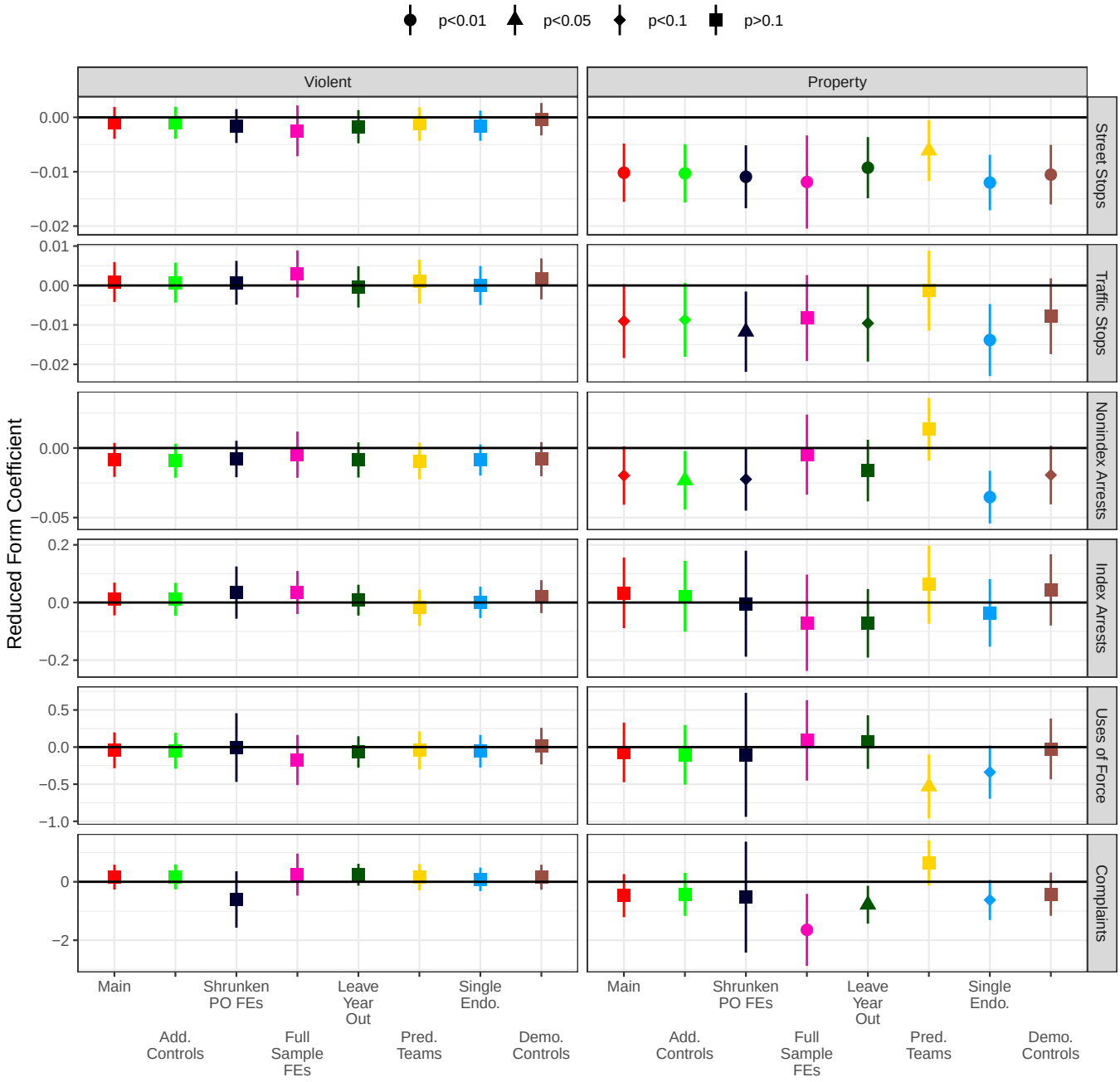
Note: Figure compares the conditional correlations computed from the implied regression (equation (4)) of regressing officer activity effects of one activity on all the other activity effects using the bias-corrected (co)variance structure. Estimates are re-scaled to be conditional correlations. Shift-level estimates are the focus of Panel A of Figure 4, and the point estimates and standard errors in text correspond to those of the team-level estimates. Labels display resulting R^2 of the implied regressions of team-level design. Bootstrap standard errors are used for 95% confidence intervals for both, with team-level bootstrap runs coinciding the draws used to compute splits.

Figure A.7: Team-Level Officers Effects on 911 Calls and Relationship to Effects on Crime



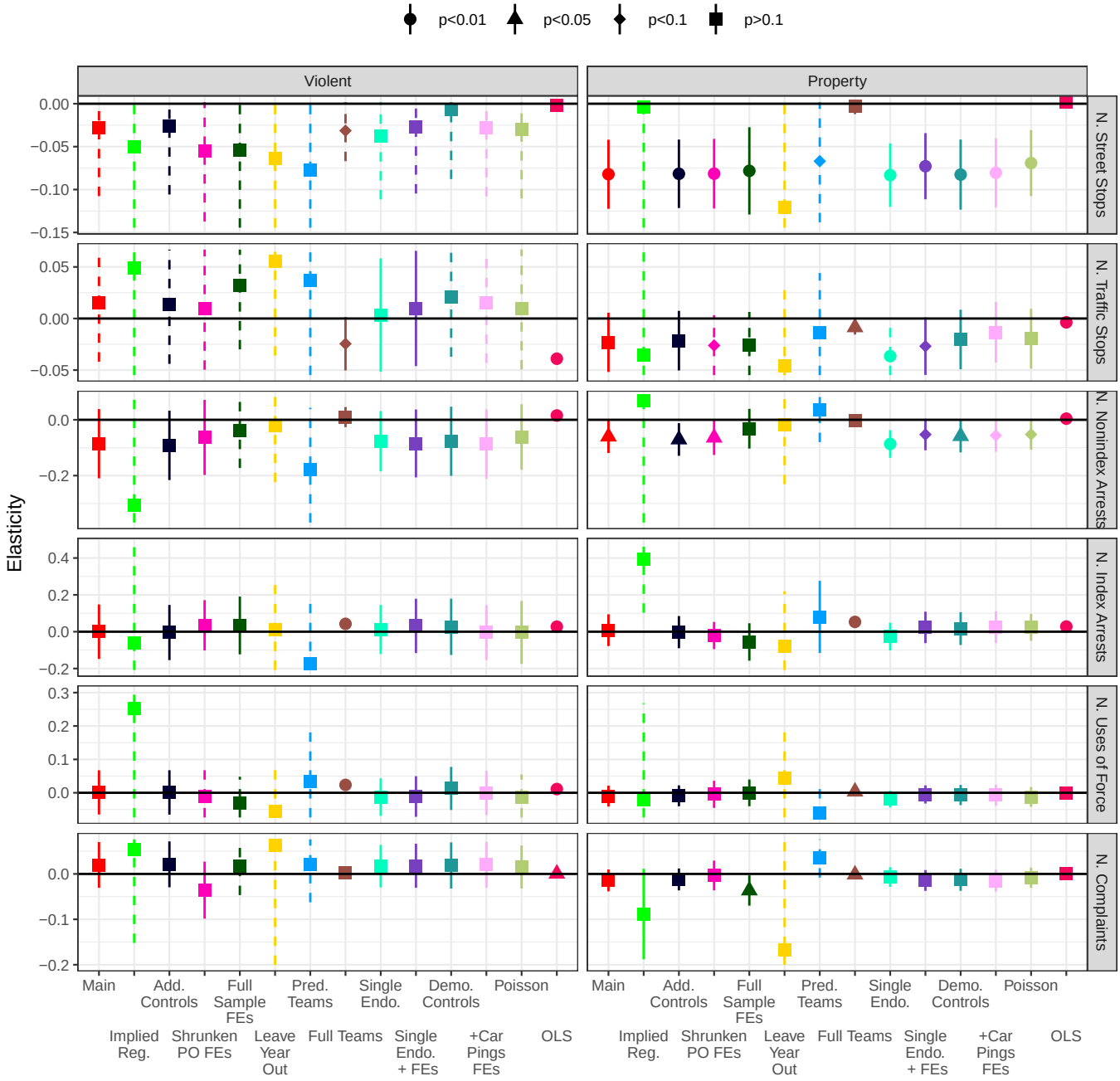
Note: Figure displays results relating to team-level officer effects on 911 calls as the outcome in equation (2). Panel A displays split-sample bias-corrected correlations between officer effects on 911 calls of various types and officer effects on violent and property crime reports. Panel B displays the variation explained by variation in officer effects for call types and violent and property crime. Bootstrap standard errors are used for 95% confidence intervals, with bootstrap runs coinciding the draws used to compute splits.

Figure A.8: Reduced Form Estimates across Robustness Specifications



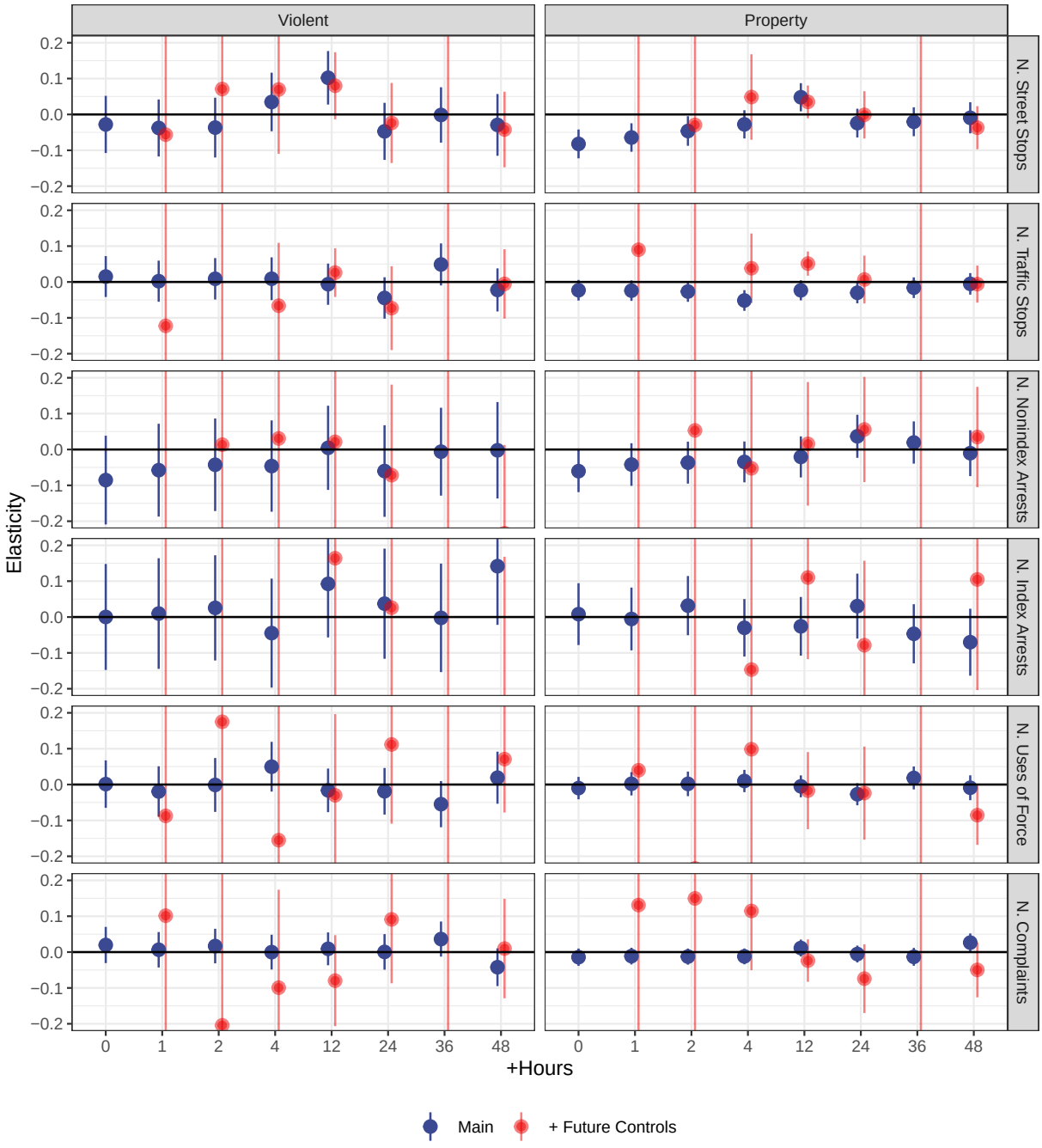
Note: Figure displays estimates of reduced form effects of team propensities for policing activities on index violent and property crime using equation (5). 'Main' refers to the main estimates (Table 5). See Appendix B.2 for details on specifications. Standard errors are clustered at the district-year-month-day-of-week-hour level and are used to construct 95% confidence intervals, and point shape denotes their statistical significance.

Figure A.9: Elasticity Estimates across Robustness Specifications



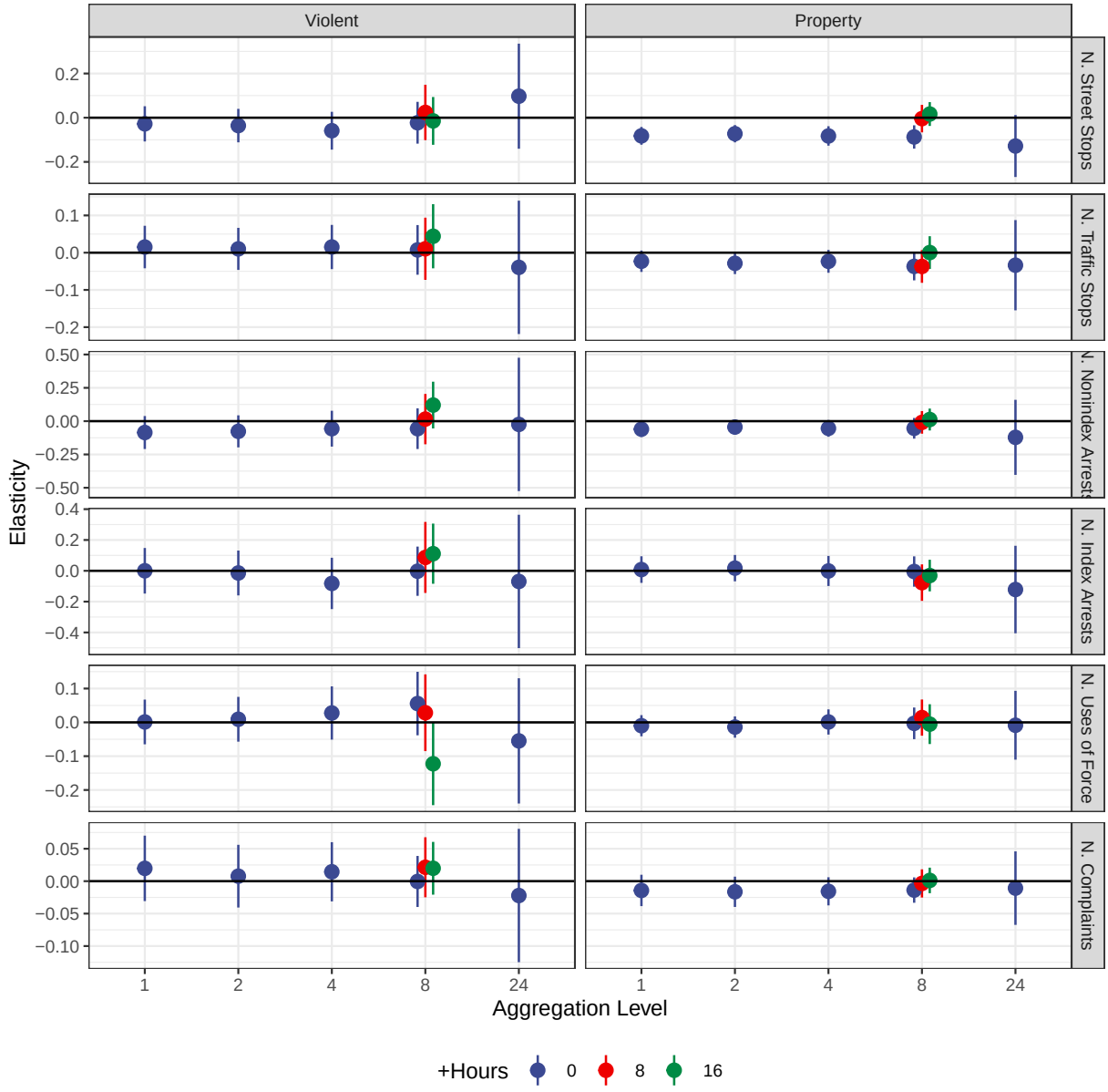
Note: Figure displays estimates of the elasticity of index violent and property crime with respect to each policing activity using 2SLS estimates from equations (6) and (7) multiplied by the mean activity and divided by the mean outcome, except for 'Poisson' which uses the control function method, along with associated 95% confidence intervals. 'Main' refers to the main estimates (Table 5), and OLS refers to equation (6). See Appendix B.2 for details on specifications. Standard errors are clustered at the district-year-month-day-of-week-hour level for all specifications other than 'Implied Reg.' for which bootstrapped standard errors are used. Standard errors used to construct 95% confidence intervals. Dashed lines denote a 95% confidence interval has been trimmed for visibility, while point shape denote their statistical significance.

Figure A.10: Elasticity of Future Crimes with Respect to Present Policing Activities



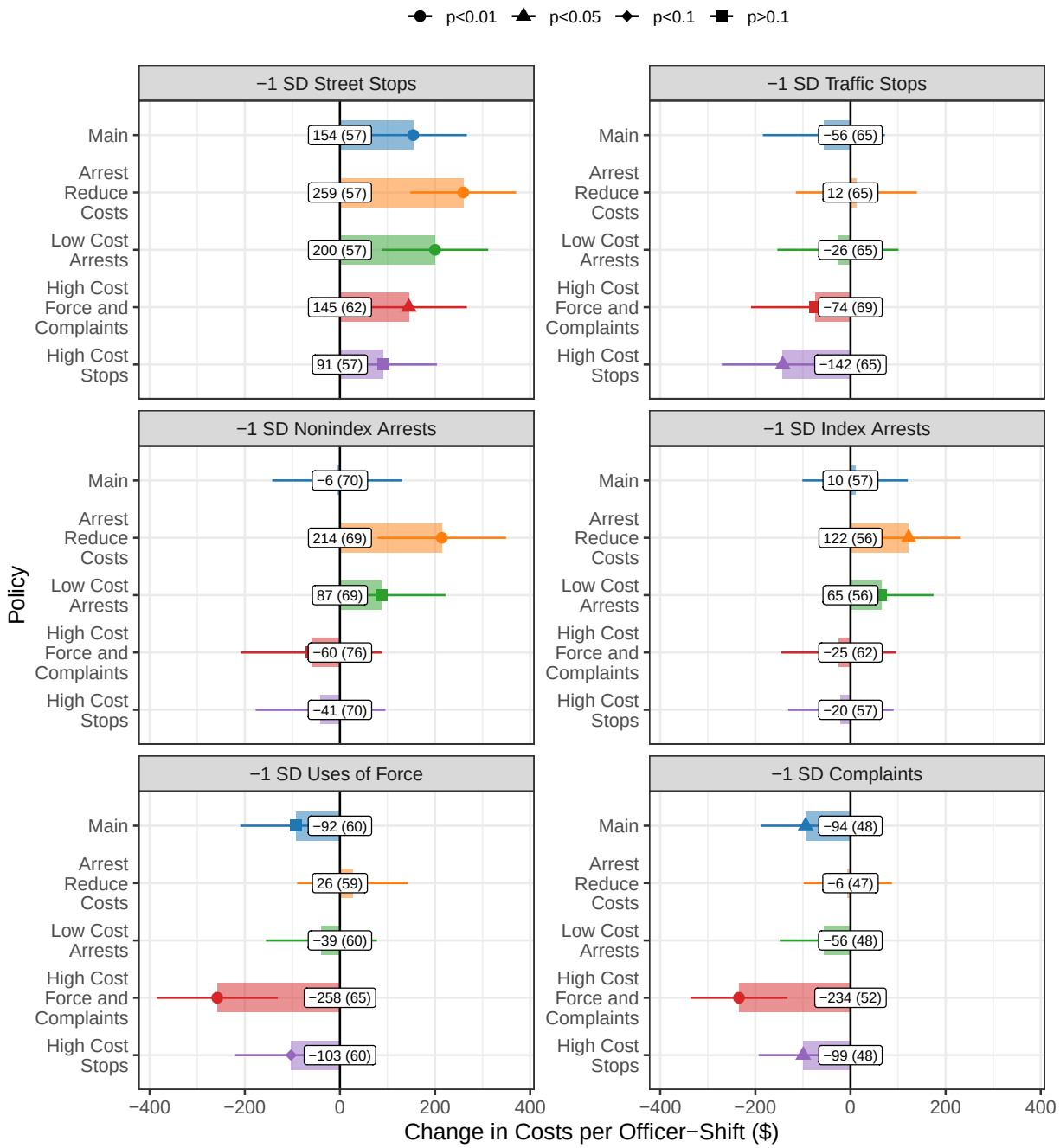
Note: Figure displays the elasticities of index violent and property crime in the same district in h hours in the future with respect to policing activities in the present, recovered by estimating equations (6) and (7) and dividing the 2SLS coefficient by the mean of the dependent variable and multiplying by the mean of the policing activity. Main refers to substituting only the future outcome in for the present outcome in equation (6). '+ Future Controls' controls for the future team composition ($\hat{\Theta}_{d,t+h}$). Standard errors are clustered at the district-year-month-day-of-week-hour level and are used to construct 95% confidence intervals. Figure is truncated for visibility.

Figure A.11: Elasticity of Crimes with Respect to Policing Activities using Alternative Aggregations



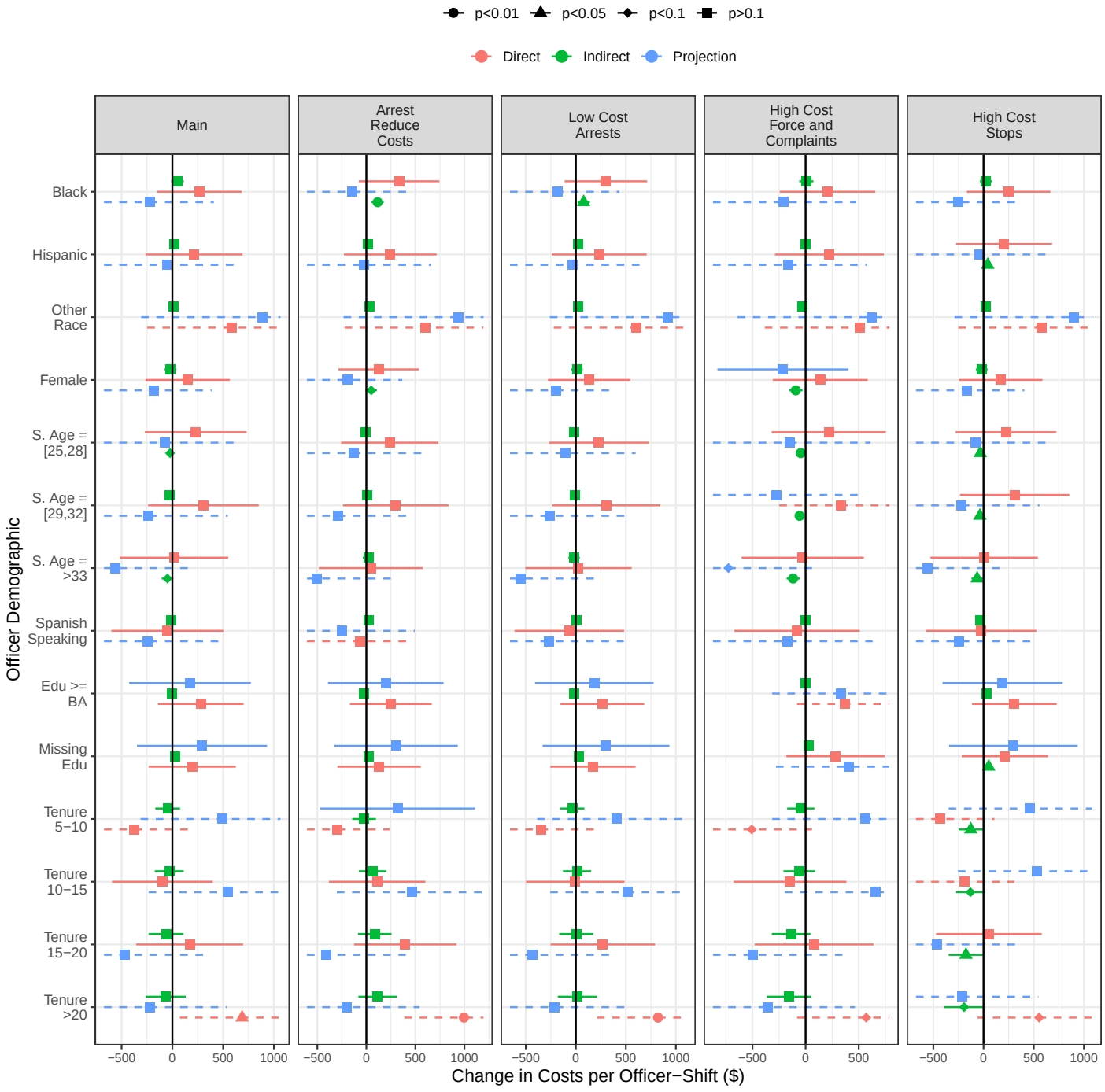
Note: Figure displays the elasticities of index violent and property crime at different levels of district-by-h hours aggregation and estimated using equations (6) and (7), where h is 1 hour in the main specification and 2, 4, 8, 12, and 24 hours in the alternative aggregations). Data is aggregated by summing over the district-by-hour data to the specified level (e.g., $\Theta^k_{d,2t}$ is the sum of $\Theta^k_{d,t}$ and $\Theta^k_{d,t+1}$) with crime being the total number of crimes in that period of time. Aggregation levels were determined by flooring the hour of day divided by h. 8 Hour Aggregation levels also include future crimes aggregated at the same levels (+ Hours 8 and 16 refer to 1 and 2 8-hour periods in the future) as in the prior figure. The corresponding fixed effects for Γ are district-year-month-day-of-week- h hour period, and standard errors are clustered at that level to construct 95% confidence intervals.

Figure A.12: Effects of Replacing Officers on Total Costs (Low Murder) per Officer-Shift under Alternative Activity Costs



Note: Figure displays the effects on total costs (low murder) of replacing an officer whose officer effect for a specific activity is 1 standard deviation above the mean with an average officer. Effects are scaled to be per officer 9-hour shift. Changes in costs take into account the 1 SD decline in the focal activity and the implied changes in other activities based on the bias-corrected covariances between officer effects across activities. Officer effects are computed from estimating equation (1), and standard deviations and covariances are KSS bias-corrected. Changes in activities are converted into changes in costs by applying changes in officer effects to reduced form coefficients of the effect of team propensities on costs at the district-hour level (equation (5)) with standard errors clustered at the district-year-month-day-of-week-hour level. Estimates use the reduced form effects of team propensities on total costs, with point estimates and corresponding standard errors computed using a linear combination of the coefficients and covariance matrix. Points and colors denote different activity costs: Main refers to the main estimates (Figure 6); Arrests Reduce Costs applies -$2,500 and -$5,000 to nonindex arrests and index arrests; Low Cost Arrests applies $2,000 and $0 to nonindex arrests and index arrests; High Cost Force and Complaints applies $100,000 and $60,000 to use of force and complaints; High Cost Stops applies $500 to street and traffic stops.

Figure A.13: Effects of Officer Demographics on Total Costs (Low Murder) per Officer-Shift



Note: Figure displays the effects of officer demographics on (low murder) total costs per 9-hour officer-shift along with 95% confidence intervals. Estimates can be interpreted as the effect of having an officer with a specific demographic on a team relative to the reference officer (white, male, started in the CPD younger than 25, do not speak Spanish, have less than a BA in education, and tenure between 1.5 and 5 years). 'Direct' regresses costs at the district-hour level onto team demographics (equation (8)), with standard errors clustered at the district-year-month-day-of-week-hour level and used to construct confidence intervals. The 'Indirect' method uses the projection coefficients from regressing shift-level officer fixed effects estimated from equation (1) on officer demographics (equation (3)) multiplied the reduce form effects of officer effects on costs (equation (5)) to construct the expected change in costs (equation (9)), with linear combinations of the variance-covariance matrix of reduce form effects used to compute standard errors. 'Projection' uses the projection coefficients recovered from regressing team-level officer effects for costs recovered from estimating equation (2) on officer demographics (equation (3)) weighted by officer-shift observations, with heteroskedasticity robust standard errors (HC1) used to compute confidence intervals. Main refers to the main costs estimates (Figure 7); Arrests Reduce Costs applies -$2,500 and -$5,000 to nonindex arrests and index arrests; Low Cost Arrests applies $2,000 and $0 to nonindex arrests and index arrests; High Cost Force and Complaints applies $100,000 and $60,000 to use of force and complaints; High Cost Stops applies $500 to street and traffic stops.

Table A.1: Elasticity of Crime Subtypes with respect to Policing Activities

	Index Violent					Index Property				Nonindex			
	All Violent	Murder	Crim. SA	Robbery	Agg. Assault	All Property	Burglary	Motor Veh. Theft	Theft	Drug	Simple Violent	Weapon	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Street Stops	-0.028	0.017	-0.77***	0.015	0.028	-0.082***	-0.11**	-0.1*	-0.073***	0.056	-0.048*	0.32**	-0.022
	(0.041)	(0.28)	(0.19)	(0.061)	(0.058)	(0.021)	(0.048)	(0.056)	(0.024)	(0.067)	(0.026)	(0.13)	(0.023)
Traffic Stops	0.015	-0.15	0.22*	0.014	-0.0033	-0.023	0.0037	0.037	-0.041**	-0.094**	0.038**	0.054	-0.0056
	(0.029)	(0.23)	(0.12)	(0.042)	(0.043)	(0.015)	(0.031)	(0.038)	(0.018)	(0.037)	(0.019)	(0.098)	(0.016)
Nonindex Arrests	-0.086	-0.57	-0.27	0.009	-0.13	-0.06**	-0.049	0.0037	-0.081**	1.3***	0.0077	0.13	0.13***
	(0.063)	(0.45)	(0.26)	(0.096)	(0.091)	(0.03)	(0.074)	(0.087)	(0.035)	(0.11)	(0.04)	(0.2)	(0.034)
Index Arrests	-0.000048	-0.47	0.014	0.088	-0.06	0.0081	0.1	-0.0099	-0.014	-0.4***	0.034	0.44*	0.011
	(0.076)	(0.52)	(0.32)	(0.12)	(0.11)	(0.044)	(0.088)	(0.1)	(0.056)	(0.13)	(0.05)	(0.24)	(0.045)
Uses of Force	0.0012	0.094	-0.082	-0.035	0.04	-0.01	-0.00061	-0.0094	-0.013	0.076	0.045**	0.27**	-0.0087
	(0.034)	(0.24)	(0.14)	(0.051)	(0.048)	(0.016)	(0.038)	(0.045)	(0.019)	(0.053)	(0.021)	(0.11)	(0.018)
Complaints	0.02	0.055	-0.027	0.0041	0.038	-0.014	-0.022	-0.018	-0.012	0.14***	-0.0039	0.25***	-0.012
	(0.026)	(0.18)	(0.11)	(0.039)	(0.037)	(0.012)	(0.03)	(0.035)	(0.014)	(0.041)	(0.016)	(0.082)	(0.014)
Mean Dep.	0.13	0.0026	0.0083	0.057	0.064	0.49	0.086	0.064	0.34	0.12	0.3	0.02	0.42

Note: Table displays the elasticities of all crime types with respect to policing activities. Elasticities are constructed using the 2SLS estimates from equations (6) and (7), multiplying the coefficient by the mean activity and dividing it by the mean outcome per district-hour. Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Reduced Form Effects of Activities on Costs ($1,000)

	Activity Costs	Crime Costs		Total Costs (Activity + Crime)	
		Full Murder	Low Murder	Full Murder	Low Murder
	(1)	(2)	(3)	(4)	(5)
Street Stops	-0.077**	-0.37	-0.91***	-0.44	-0.98***
	(0.03)	(1.6)	(0.22)	(1.6)	(0.22)
Traffic Stops	0.062	-2	0.45	-1.9	0.51
	(0.043)	(3.2)	(0.36)	(3.2)	(0.37)
Nonindex Arrests	1.7***	-8.3	-0.94	-6.7	0.72
	(0.13)	(6.8)	(0.84)	(6.8)	(0.86)
Index Arrests	-0.13	-18	1.1	-18	0.93
	(0.63)	(30)	(4)	(30)	(4.1)
Uses of Force	18***	17	7.3	35	26
	(2.4)	(140)	(17)	(140)	(17)
Complaints	15***	25	1.9	40	17
	(4.3)	(240)	(30)	(240)	(31)
Mean Dep	3.48	41.53	22.1	45.02	25.59

Note: Table displays coefficients of the reduced form relationship between activity, crime, and total costs at the district-hour level and team activitiy effects using equation (5). Coefficients can be interpreted as the effect of a 1 unit increase in total team effect for that specific activity effect. Crime costs refer to total crime costs using Miller et al. (2021) incidence costs applied to reported crimes and 'Low Murder' refers to using a value of $400,000 rather than $8 million ('Full Murder'). Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Effects of Replacing Officers on Costs

	-1 SD in θ_i^k					
	Street Stops	Traffic Stops	Nonindex Arrests	Index Arrests	Use of Force	Complaints
	(1)	(2)	(3)	(4)	(5)	(6)
Activity Cost						
All Coefs	-8.2***	-5.7***	-18***	-9.5***	-14***	-11***
	(0.87)	(0.84)	(1.1)	(0.91)	(0.89)	(0.72)
Correct Sign	-10	-6.4	-19	-11	-15	-11
Crime Cost - Full Murder						
All Coefs	71	65	102.1*	72	36	21
	(49)	(62)	(62)	(49)	(53)	(41)
Correct Sign	73	67	105.7	75	46	29
Crime Cost - Low Murder						
All Coefs	25***	-0.49	17**	11*	4.2	0.67
	(6.3)	(7.2)	(7.6)	(6.2)	(6.6)	(5.2)
Correct Sign	29	9.2	22	15	9.1	4
Total Cost - Full Murder						
All Coefs	63	60	84	63	22	10
	(49)	(62)	(62)	(49)	(53)	(41)
Correct Sign	62	60	86	65	31	18
Total Cost - Low Murder						
All Coefs	17***	-6.2	-0.66	1.1	-10	-10**
	(6.4)	(7.3)	(7.7)	(6.3)	(6.7)	(5.3)
Correct Sign	19	2.9	2.3	4.2	-5.9	-7.3

Note: Table displays effects of replacing and training officers on costs applied to one officer over the course of a 9-hour officer-shift. -1 SD refers to replacing an officer with + 1 SD in terms of one type of officer effect with an average officer. Crime costs refer to total crime costs using Miller et al. (2021) incidence costs applied to reported crimes and 'Low Murder' refers to using a value of $400,000 rather than $8 million ('Full Murder'). 'Correct Sign' refers to constructing effects excluding reduced form coefficients which suggest that officer activity reduces activity costs or increases crime costs to activity and crime costs, then summing over them to construct total costs (no standard errors are computed). Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses and constructed using a linear combination over the variance-covariance matrix from the reduced form regressions. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Effects of Officer Demographics on Costs ($1,000 per 900 Hours)

	Black	Hispanic	Other Race	Female	S. Age = [25,28]	S. Age = [29,32]	S. Age = >33	Spanish Speaking	Edu >= BA	Missing Edu	Tenure 5-10	Tenure 10-15	Tenure 15-20	Tenure >20
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Activity Cost														
Direct	-56**	-18	-36	12	-6.1	14	-28	4.6	45*	64**	-84***	-140***	-160***	-220***
	(26)	(28)	(49)	(25)	(29)	(32)	(32)	(33)	(26)	(26)	(31)	(29)	(30)	(34)
Indirect	-5.7***	-0.45*	-2.8***	-6.8***	-1.8***	-3.1***	-6.7***	-2***	1.6***	0.45*	-1.6**	-7.3***	-12***	-15***
	(0.42)	(0.26)	(0.29)	(0.4)	(0.17)	(0.21)	(0.38)	(0.25)	(0.27)	(0.25)	(0.72)	(0.85)	(1)	(1.2)
Proj.	-4.3	-5.2	-11	0.027	2.1	3	-8.1	2.3	3.2	2.5	13**	8.7	-5.3	-5.6
	(4.6)	(4.6)	(8)	(4.1)	(5)	(5.4)	(5.4)	(5.3)	(4.1)	(4.4)	(5.3)	(5.8)	(5.8)	(5.7)
Crime Cost - Full Murder														
Direct	3027*	1483	10396***	-510	2010	1913	-2800	1291	-800	-1200	-610	2641	2541	2958
	(1586)	(1796)	(2898)	(1536)	(1799)	(1926)	(1939)	(1972)	(1608)	(1575)	(2033)	(1931)	(1972)	(2225)
Indirect	29	-14	-3.5	17	8.9	15	27	23	-22	-10	42	76	98	116
	(24)	(17)	(18)	(23)	(11)	(13)	(23)	(17)	(19)	(17)	(54)	(62)	(74)	(84)
Proj.	20	-110	1103**	-170	-160	-260	-690**	634**	-200	-490**	-48	452	237	48
	(256)	(282)	(446)	(226)	(284)	(309)	(311)	(305)	(244)	(250)	(310)	(318)	(303)	(297)
Crime Cost - Low Murder														
Direct	296	210	563	124	213	261	42	-49	206	113	-260	48	314	838***
	(188)	(216)	(377)	(188)	(227)	(246)	(242)	(249)	(191)	(194)	(245)	(224)	(238)	(278)
Indirect	11***	2.6	3.9*	5*	-0.48	0.7	1.9	1.1	-2.2	2.5	-3	4.2	6	8.3
	(3.1)	(2.1)	(2.2)	(2.9)	(1.3)	(1.6)	(2.9)	(2)	(2.2)	(2)	(6.1)	(7.2)	(8.7)	(9.9)
Proj.	-18	0.17	100*	-18	-9.5	-26	-48	-27	14	27	36	46	-42	-17
	(31)	(35)	(60)	(28)	(35)	(39)	(39)	(38)	(30)	(32)	(40)	(39)	(41)	(38)
Total Cost - Full Murder														
Direct	2971*	1465	10360***	-500	2004	1927	-2800	1296	-760	-1200	-700	2504	2381	2739
	(1586)	(1796)	(2899)	(1537)	(1799)	(1927)	(1940)	(1972)	(1608)	(1576)	(2033)	(1931)	(1972)	(2226)
Indirect	23	-14	-6.3	11	7.1	12	20	21	-20	-9.7	41	69	86	101
	(24)	(17)	(18)	(23)	(11)	(13)	(23)	(17)	(19)	(17)	(54)	(62)	(74)	(84)
Proj.	16	-120	1092**	-170	-150	-260	-700**	636**	-190	-490*	-35	461	232	42
	(257)	(282)	(445)	(226)	(284)	(310)	(311)	(305)	(244)	(250)	(311)	(318)	(303)	(298)
Total Cost - Low Murder														
Direct	240	192	527	136	207	275	14	-44	252	176	-340	-89	154	618**
	(191)	(219)	(383)	(191)	(230)	(250)	(246)	(253)	(194)	(197)	(248)	(228)	(242)	(282)
Indirect	5	2.2	1.1	-1.8	-2.3*	-2.4	-4.8*	-0.83	-0.56	3	-4.6	-3.1	-6.2	-6.5
	(3.2)	(2.1)	(2.3)	(3)	(1.3)	(1.6)	(2.9)	(2)	(2.3)	(2)	(6.2)	(7.3)	(8.8)	(10)
Proj.	-22	-5	89	-18	-7.4	-23	-56	-25	17	29	49	55	-47	-22
	(32)	(35)	(61)	(29)	(36)	(40)	(39)	(38)	(31)	(33)	(41)	(40)	(42)	(39)

Note: Table displays the effects of an officer demographic, controlling for all others, on costs over 900 officer-hours, with standard errors in parentheses. Estimates can be interpreted as the effect of having an officer with a specific demographic on a team relative to the reference officer (white, male, started in the CPD younger than 25, do not speak Spanish, have less than a BA in education, and tenure between 1.5 and 5 years).'Direct' regresses costs at the district-hour level onto team demographics (equation (8)), with standard errors clustered at the district-year-month-day-of-week-hour level and used to construct confidence intervals. The 'Indirect' method uses the projection coefficients from regressing shift-level officer fixed effects estimated from equation (1) on officer demographics (equation (3)) multiplied the reduce form effects of officer effects on costs (equation (5)) to construct the expected change in costs (equation (9)), with linear combinations of the variance-covariance matrix of reduce form effects used to compute standard errors. The 'Indirect $+$ Correct Sign Coef.' method only uses the correct signed coefficients (coefficients which have activities weakly increasing activity costs and weakly decreasing crime costs) for the Indirect method. 'Projection' uses the projection coefficients recovered from regressing team-level officer effects for costs recovered from estimating equation (2) on officer demographics (equation (3)) weighted by officer-shift observations, with heteroskedasticity robust standard errors (HC1) used to compute confidence intervals. Crime costs refer to total crime costs using Miller et al. (2021) incidence costs applied to reported crimes and 'Low Murder' refers to using a value of $400,000 rather than $8 million ('Full Murder'). *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Effects of Officer Demographics on Crimes per 900 Hours

	Index Violent		Index Property		Nonindex Drug		Nonindex Simple Violent		Nonindex Weapon		Nonindex All Other	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Black	1.1	-6.7	-2.6	-13	-6.1***	-12**	4.9**	-14*	-1.3**	-1.3	1.4	-2
	(1.4)	(5.2)	(2.5)	(9)	(1.4)	(5.3)	(2)	(7.5)	(0.54)	(2)	(2.4)	(8.7)
Hispanic	1	-12*	-1.3	-17	3.6**	11*	0.77	-5.1	-0.24	0.54	-0.7	-8.1
	(1.5)	(6.2)	(2.8)	(11)	(1.5)	(6.3)	(2.3)	(8.8)	(0.62)	(2.4)	(2.7)	(10)
Other Race	0.74	-26***	-6.9	-25	2.5	25**	2	-17	-1.5	1.4	2.3	3.4
	(2.6)	(9.8)	(5.2)	(18)	(2.5)	(9.9)	(3.9)	(14)	(1)	(3.7)	(4.8)	(17)
Female	1.2	7.7*	9.2***	11	-2.1	-14***	4.5**	4.5	-0.98*	-0.42	2	3
	(1.4)	(4.4)	(2.6)	(7.6)	(1.3)	(4.5)	(2)	(6.3)	(0.54)	(1.7)	(2.4)	(7.4)
S. Age = [25,28]	0.79	9.7	-1.4	5.4	-3.9**	-19***	2.6	11	0.088	0.81	2.2	5.2
	(1.6)	(7.1)	(3)	(13)	(1.6)	(7.3)	(2.3)	(10)	(0.63)	(2.7)	(2.8)	(12)
S. Age = [29,32]	1.6	13*	-2.7	6.6	-3.1*	-21***	0.91	15	0.32	1.9	0.83	-1.4
	(1.7)	(7.7)	(3.3)	(14)	(1.8)	(8.1)	(2.6)	(11)	(0.69)	(3)	(3.1)	(13)
S. Age = >33	0.69	0.027	1.2	5.5	-5.2***	1.5	-1.2	-7.4	-1.1	1.2	0.56	4.8
	(1.7)	(4.5)	(3.3)	(7.9)	(1.7)	(4.6)	(2.6)	(6.4)	(0.69)	(1.7)	(3.1)	(7.5)
Spanish Speaking	0.98	3.3	-3.6	14*	-3.3*	-8.4*	1.1	8.6	-1.4*	-1.2	0.87	4.8
	(1.8)	(4.4)	(3.5)	(8)	(1.8)	(4.5)	(2.7)	(6.3)	(0.7)	(1.6)	(3.2)	(7.5)
Edu >= BA	1.6	-8.7	6.1**	-2.3	-1.7	14*	1.1	-0.035	-0.69	-1.1	9.5***	14
	(1.4)	(7.2)	(2.5)	(12)	(1.4)	(7.4)	(2)	(10)	(0.56)	(2.8)	(2.5)	(12)
Missing Edu	2	9.9**	-2.2	-2.2	3.6**	-1.3	-2.7	5.4	0.18	-1.7	8.2***	14*
	(1.4)	(4.6)	(2.8)	(8.2)	(1.4)	(4.7)	(2.1)	(6.6)	(0.57)	(1.8)	(2.6)	(7.9)
Tenure 5-10	-0.36	33**	1.6	21	-2	-39**	-7.7***	23	-1.2*	-4.1	0.21	20
	(1.7)	(15)	(2.8)	(26)	(1.7)	(15)	(2.4)	(22)	(0.71)	(6.1)	(2.9)	(26)
Tenure 10-15	2.4	43**	4	35	-9.1***	-58***	-1.9	34	-1.4**	-2.9	-5.8**	9.6
	(1.6)	(20)	(2.7)	(33)	(1.5)	(20)	(2.3)	(28)	(0.67)	(7.7)	(2.7)	(33)
Tenure 15-20	2.9*	56**	15***	43	-8.6***	-71**	-0.56	49	-2.1***	-5.7	1.3	25
	(1.7)	(27)	(3.1)	(45)	(1.6)	(28)	(2.4)	(38)	(0.69)	(11)	(2.9)	(45)
Tenure >20	2.9	54**	22***	63	-12***	-74***	5.4*	56	-2.1***	-6.5	-5.6	5.9
	(1.9)	(27)	(3.8)	(45)	(1.8)	(28)	(2.9)	(38)	(0.8)	(11)	(3.5)	(45)

Note: Table displays the effects of an officer demographic, controlling for all others, on costs over 900 officer-hours. Estimates can be interpreted as the effect of having an officer with a specific demographic on a team relative to the reference officer (white, male, started in the CPD younger than 25, do not speak Spanish, have less than a BA in education, and tenure between 1.5 and 5 years). OLS corresponds to the 'Direct' method which regresses costs at the district-hour level onto team demographics (equation (8)), with standard errors clustered at the district-year-month-day-of-week-hour level and used to construct confidence intervals. IV estimates use the predicted composition of the team based on the rotation calendar prediction as instruments for the actual composition. Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B Additional Analyses

B.1 Robustness Specifications for Officer Effects

In the first robustness specification, I separately estimate equation (1) within each of the three main sample periods (2010-2012, 2013-2015, and 2016-2019) to ensure the results are not driven by the specific sample period used ("Periods").

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, which was a concern due to the presence of pre-trends ("Donut"). Additionally, I take the sample of observations in which the officer is predicted to be working based on the rotation 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"). Note that while the no partner specification does not drop a large number of officer effects, it does significantly reduce the number of observations per officer because having any partner

depends more on being in a specific watch rather than a specific unit.

To ensure the results are consistent when using the same filters as the district-hour analysis, I use only patrol units and regular watches 1-4 ("Regular Units/Watches"). To test robustness to less granular assignment definitions which do not change as frequently throughout an officer-year, I define an assignment as a unit-watch ("Assign = Unit-Watch").

Finally, I aggregate the data to the officer-week level and select the officer's most common assignment within that week 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, and serial correlation is reduced.

To summarize, the results of the robustness tests are consistent with the main results. Figure A.2 tests officer effects' stability across these different samples by taking the correlation between officer effects from the main sample and each alternative sample/specification. These correlations are not bias-corrected and thus are biased downwards. Nevertheless, the cross-sample correlations are very strong, generally > 0.75 for most outcomes, with the lowest being complaints and use of force and results from small subsamples. Figure A.3 shows that standard deviations of officer effects (scaled by the mean of the outcome) are generally consistent across specifications. Figure A.4 shows that cross-outcome correlations in officer effects are also similar across specifications. If anything, they are generally more negative than the main results, which reinforces the weak relationship between use of force and misconduct and other activities. Naturally, differences are most noticeable in samples that deviate significantly from the main sample.

B.2 Robustness Specifications for Crime Effects

This appendix first summarizes the additional and robustness specifications used for the 2SLS results and the reduced form (RF) results discussed in Section 4 and displayed in Figures A.8 and A.9, with 2SLS estimates re-scaled to be elasticities for ease of comparability. Note that not all specifications can be applied to the reduced form analysis. I begin discussing the tests

relating to exclusion restriction violations, then discuss all other tests.

Violations of the exclusion restriction (that team propensities only influence crime through influencing the activities themselves) is a common concern in the judge-design literature. Under constant treatment effects, exclusion violations produced through observable activities (e.g., teams which make more stops also make more arrests, which can influence crime) can be controlled for using the team's propensity for the other activities (Mueller-Smith (2015), Kamat et al. (2023), Humphries et al. (2023)). Under this assumption, I perform multiple robustness tests (see Appendix B.3 for interpretation under heterogeneous treatment effects). The results are consistent across each of these tests.

'Single Endo.' only uses one endogenous variable at a time, showing the sensitivity of results to using multiple endogenous variables. The main results are unaffected, other than use of force having a larger negative effect on property crime, consistent with its correlation with nonindex arrests and, to a lesser extent, street stops.

'Single Endo. + FEs' uses one endogenous variable (k) and adds $\Theta^{l}_{d,t} \forall l \neq k$ as controls. The main results are unaffected.

'Demo. Controls' controls for team demographic composition by adding $Z_{d,t} = \sum_{i \in \tau(d,t)} z_i$ as controls, where z_i is a vector of officer demographics. These include dummy variables indicating: race (Black, Hispanic, other non-white), gender (female), start age ($25-28$, $29-32$, ≥ 33), tenure ($5-10$,$10-15$,$15-20$, >20). This ensures that the results are not driven by unobserved behaviors correlated with both policing activities and activity effects and also team demographics. The main results are unaffected.

'+ Car Ping FEs' controls for 'car ping' fixed effects. I link data on police car GPS pings from 2015 - 2019, I estimate officer fixed effects for both average car speed and number of pings during their shifts using equation (1). Then I take the estimated officer fixed effects for ping speed and quantity to construct team propensities for both of these behaviors. Then I add team propensities ($\Theta^{k}_{d,t}$s) for speed and ping quantity as controls in the 2SLS and RF specifications. For times outside of 2015-2019, the control values are all 0. While this is not

a perfect proxy for officer activity levels, if the frequency of pings and speed at which officers drive is either positively or negatively correlated with their underlying activity level, then including this control should alter estimates. In practice, it has little effect, suggesting either it is a poor proxy for activity levels or activity levels beyond observed policing activities are not generating significant violations of the exclusion restriction.

'Implied Reg.' displays implied regression coefficients (equation (4)) of regressing ω_i^c on a vector of ω_i^ks for all activity effects, but using the split-sample bias-corrected (co)variance structure to estimate the coefficients.

'Add Controls' adds controls for the number of officers reported working across ranks (detectives, sergeants, etc.) and duty types (desk duty, overtime, special officers, etc.) in addition to the total number of officers on the team. The main results are unaffected.

'Shrunken PO FEs' uses officer fixed effects shrunken using the Empirical Bayes method from Weisburst (2022) to construct $\Theta_{d,t}$, reducing measurement error to avoid attenuating in the RF results and reducing noise in the 2SLS results (Frandsen et al. (2025)). The main results are unaffected.

'Full Sample FEs' uses officer fixed effects computed using the entire sample from 2010-2019 (as in Section 3) rather than separate estimation for each sample period. While this produces a slightly weaker first stage and noisier RF, it ensures that using sample period specific officer effects to construct $\Theta_{d,t}$ does not bias the results.

'Leave Year Out' computes fixed effects for the first stage excluding the focal year from the sample. Again, this reduces the first stage and RF strength, but it avoids any concerns that contemporaneous shocks to crime and policing conditions are also included in the estimation of θ_i, and thus $\Theta_{d,t}$. This is similar to a jackknife IV procedure.

'Poisson' uses a poisson model and the control function method to correct for endogeneity (Wooldridge (2015)). The coefficients and standard errors are transformed by multiplying by the mean of the variable of interest. The results are consistent with those of the linear model.

'Pred. Teams' specification uses only variation in who is working based on the predicted

rotation, which removes any influence of short-run changes in crime conditions which would bias estimates. This reduces first stage and RF strength, since it mis-measures the exact set of officers working and thus $\Theta_{d,t}$, but it ensures instrumented team composition is completely independent of short-run shocks to policing and crime. This also takes care of another concern: contemporaneous policing generates feedback into contemporaneous shocks or that idiosyncratic shocks generate surges in policing activity and change the set of officers working by pulling from other locations. As this would be determined by short-run shocks, the predicted set of officers working based on the calendar rotation are not susceptible to it. Nevertheless, because the predicted teams are become a weak instrument for actual teams as the year progresses (see Figure 1), it also becomes a weaker instrument for activities. The violent crime effects move closer to zero, while for property crimes the effect of street stops becomes slightly smaller, nonindex arrests have a positive ($p > 0.1$) effect, and use of force has a negative effect ($p < 0.05$). While not shown, filtering the data to remove later months within each year reduces sample size but also increases the instruments' strength and moves point estimates toward the main results.

'Full Teams' uses the exact composition of the team (officer dummies) at the district-hour level as instruments for policing activity, adapting equation (2), with corresponding second and first stages:

$$\text{Crime}_{d,t} = \sum_k \eta_k Y^k_{d,t} + \pi N_{d,t} + \Gamma_{d,t} + \varepsilon_{d,t} \tag{10}$$

$$Y^k_{d,t} = \sum_i \omega^k_i \times 1[i \in \tau(d,t)] + \Gamma^k_{d,t} + \epsilon^k_{d,t} \tag{11}$$

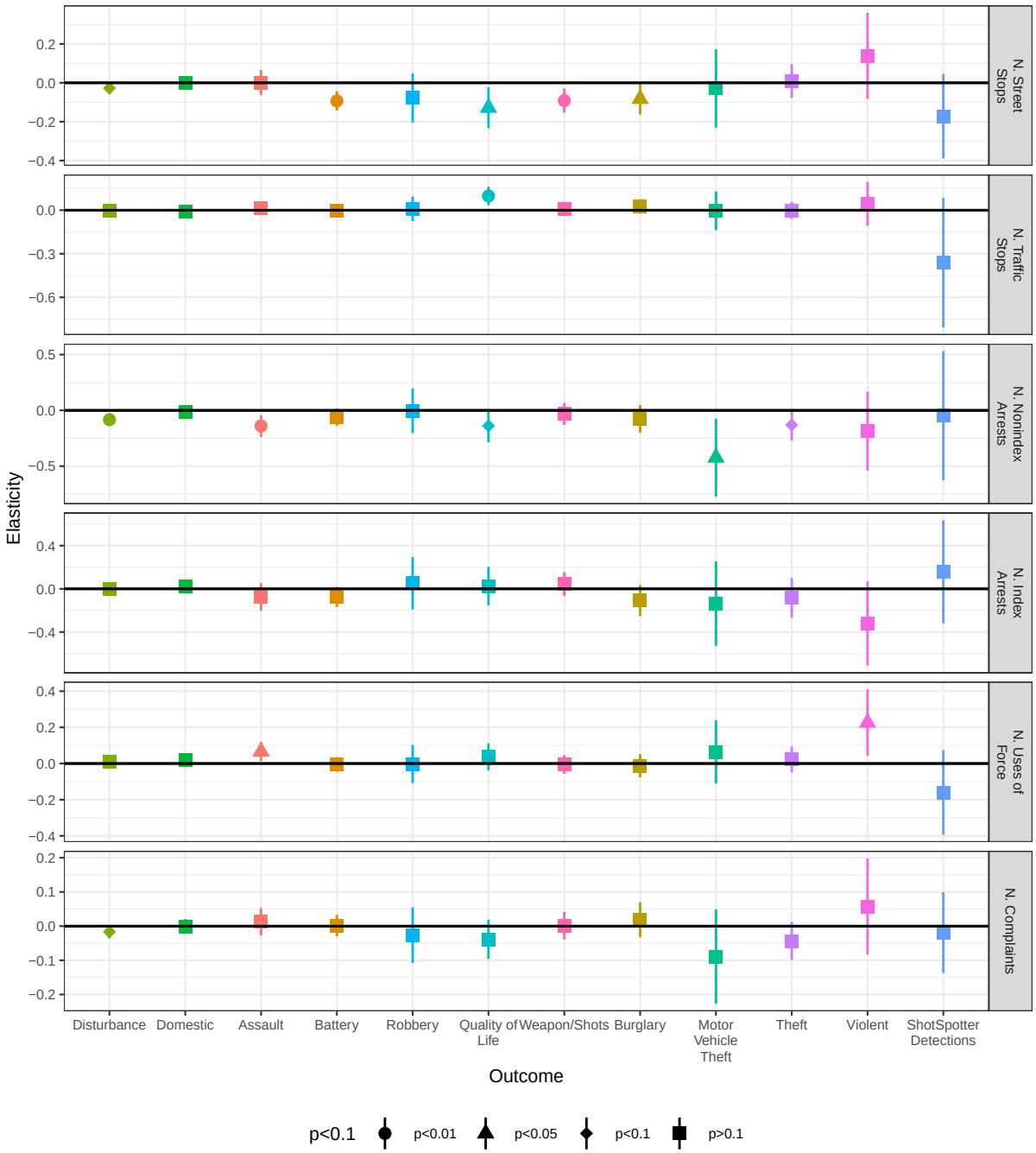
This amounts to six first stage equations with over 10,000 instruments (one for each officer). Exact officers at the district-hour level prove to be relatively weak instruments for policing activities, with all F-statistics below 10, though they are modestly sized (> 4) for the more frequent activities. This is not only due to the very large number of instruments,

but also because the district-hour model effectively mis-attributes activities to officers and attenuates individual contributions and it has significantly fewer effective observations for each officer. Standard errors are clustered at the district-year-month-day-of-week-hour level (Γ) using the proper degrees of freedom correction (Ding (2021)), and the resulting 2SLS estimates tend to be smaller in magnitude, closer to endogenous OLS, and overly precise, features consistent with known biases from using many weak instruments (Keane & Neal (2022)).

Finally, to test for spurious effect of policing on crime or the possibility that officers are influencing reporting not crime, I use 911 calls and ShotSpotter detections rather than crimes as the outcomes of interest. While 911 call types do not map directly to resulting crime reports (e.g., a call about a disturbance could be a theft, an assault, or a domestic incident), if the main results are not due to mismeasurement, then street stops and nonindex arrests should have consistent effects on related call types. While ShotSpotter detections provide a separate measure of shootings from civilian reporting (Ang et al. (2025)), the sample period is relatively small, as ShotSpotter had rolled out across Chicago by May 2018.

Figure B.1 shows the results with estimates converted into elasticities for visibility. Street stops have statistically significant elasticities on calls for battery, weapon offenses, quality of life, and burglary ranging from -0.08 to -0.13 (all $p < 0.05$) and an elasticity with disturbances of -0.028 ($p < 0.1$), while out of over 20 911 call types, traffic stops, index arrests, complaints, and uses of force collectively have few total marginally significant effects. For other activities, only a few coefficients are statistically significant, either due to spurious correlations or actual effects (e.g., uses of force increase OEMC calls about violent incidents and assaults). No activities have statistically significant effects on ShotSpotter detections, consistent with their insignificant impact on violent crime.

Figure B.1: Elasticities of 911 Call and ShotSpotter Detections with Respect to Policing Activities



Note: Figure displays the elasticities of 911 call types and ShotSpotter detections with respect to policing activities recovered by estimating equations (6) and (7) with the alternative outcomes and re-scaling the coefficients. Standard errors are clustered at the district-year-month-day-of-week-hour level to construct 95% confidence intervals.

B.3 Exclusion and Multiple Endogenous Variables

If treatment effects are heterogeneous, an additional average monotonicity assumption is required for the 2SLS estimates to be interpreted as a weighted average of treatment effects. Bhuller & Sigstad (2024) provides a test using the first stage results across different subsamples of data to test for violations of average monotonicity, similar to that of Frandsen et al. (2023). I perform this test by estimating the first stages on subsamples of individual units, days of week, years, and seasons (quarters). Consistent with average monotonicity being satisfied, Table B.1 shows that there are no subsamples in which the focal instrument has a negative and statistically significant relationship with the focal endogenous variable (e.g., effect of team effects on street stops on district-hour street stops) and, after correcting for multiple hypotheses using Holm (1979) only 1.1% of the 1470 coefficients for the effect of the non-focal instruments on focal endogenous variables (e.g., effect of team effects on street stops on district-hour uses of force) are statistically significant.

Table B.1: Violations of Average Monotonicity

	N Coefficients	Share p<0.05	Share q<0.05	Share Neg. and p<0.05
	(1)	(2)	(3)	(4)
Focal First Stage				
Total	294	0.752	0.486	0.000
District	150	0.613	0.380	0.000
Year	60	0.883	0.533	0.000
Season (Quarter)	24	0.917	0.708	0.000
Day of Week	42	0.881	0.595	0.000
Non-Focal First Stage (Cross-Effects)				
Total	1,470	0.154	0.011	0.090
District	750	0.136	0.009	0.073
Year	300	0.187	0.010	0.117
Season (Quarter)	120	0.208	0.017	0.133
Day of Week	210	0.105	0.005	0.086

Note: Table summarizes results of testing for violations of average monotonicity based on the test from Bhuller and Sigstad (2024). Test involves running the first stage (equation (7)) on subsamples as determined by the District, Year, Season (quarter of year), and Day of Week, with standard errors clustered at the district-year-month-day-of-week-hour level. The first stage coefficients are either focal or non-focal cross-effects. The results are summarized across all subsamples (Total) or by subsample. N. Coefficients denotes the number of coefficients across subsample runs that are either focal (one per outcome per subsample) or non-focal (four or five per outcome per subsample). Column (2) denotes the share of coefficients statistically significant at 5 percent significance. Column (3) denotes the share significant at 5 percent using Holm-adjusted q-values. Column (4) displays the share which are negative and have p-value less than 0.05.

B.4 Officer Quantity

I recover the effect of officer quantity (team size) on outcomes by estimating with 2SLS the following second and first stage models using the log of predicted team size (based on the rotation calendar prediction) as an instrument for the log of actual team size to remove any concerns over endogeneity (first stage F-stat $> 1,000$):

$$\mathrm{Y}^{c}_{d,t} = \alpha^{c} Log(N)_{d,t} + X'_{d,t}\gamma^{c} + \Gamma^{c}_{d,t} + \varepsilon^{c}_{d,t} \tag{12}$$

$$Log(N)_{d,t} = \rho Log(N)^{\text{Pred.}}_{d,t} + \nu_2 X_{d,t} + \Gamma_{d,t} + \epsilon_{d,t} \tag{13}$$

Table B.2 displays the implied elasticities of crime, costs, and activities with respect to officer quantity in Column (1), along with robustness specifications in Columns (2)-(6). Index crime effects are not statistically significant, though they cannot rule out most of the corresponding elasticities in Chalfin & McCrary (2018). For example, elasticities for violent crimes are generally more negative than for property crime, and murder has the largest negative elasticity (-0.65). Additionally, the elasticity of crime costs with respect to police quantity (-0.408 for FM ($p > 0.1$)) is close to that of Chalfin & McCrary (2018) (-0.473), and this is within the confidence interval of the LM crime cost elasticity. However, elasticities are positive and significant for all nonindex offenses, excluding simple violent crimes, which are more discretionary on the part of the officer.

Consistent with the positive effects of police quantity on arrests from Chalfin, Hansen, Weisburst, et al. (2022) and Weisburd (2021), more police also means more intense policing in the form of street (0.97) and traffic (0.76) stops and nonindex (0.65) and index (0.39) arrests (all $p < 0.01$) and use of force (0.63 ($p < 0.1$)). As a result, activity costs increase with team size, with an elasticity of 0.55 ($p < 0.01$). The elasticity of total costs with respect to officer quantity is -0.33 for FM and $+0.01$ for LM, though both are $p > 0.1$ and cannot reject elasticities of ± 0.2 due to imprecise crime cost estimates. Nevertheless, the clear positive effect on activity costs does raise questions over the trade-off between officer quantity and social costs.

Table B.2: Elasticities of Crime, Activities, and Costs with respect to Police Quantity

	Main (2SLS with Log(N. POs))	2SLS with N. POs	OLS with Log(N. POs)	OLS with N. POs	Poisson CF with Log(N. POs)	Poisson CF with N. POs
	(1)	(2)	(3)	(4)	(5)	(6)
Index Crimes						
Index Violent	-0.042 (0.14)	-0.125 (0.11)	0.064*** (0.021)	0.058** (0.026)	-0.001 (0.15)	-0.082 (0.11)
Murder	-0.65 (1.1)	0.345 (0.82)	-0.039 (0.16)	0.001 (0.19)	-0.97 (1.3)	0.788 (1)
Crim. SA	-0.382 (0.79)	-0.657 (0.55)	0.26*** (0.096)	0.329*** (0.11)	0.187 (0.58)	-0.312 (0.46)
Robbery	0.067 (0.2)	0.015 (0.17)	0.026 (0.032)	0.001 (0.038)	0.119 (0.24)	0.034 (0.16)
Agg. Assault	-0.07 (0.21)	-0.2 (0.16)	0.077** (0.03)	0.076** (0.036)	-0.012 (0.2)	-0.156 (0.15)
Index Property	0.068 (0.066)	0.086 (0.056)	0.101*** (0.011)	0.117*** (0.013)	0.049 (0.084)	0.071 (0.054)
Burglary	0.116 (0.16)	0.164 (0.13)	0.117*** (0.026)	0.092*** (0.029)	0.11 (0.2)	0.16 (0.13)
Motor Vehicle Theft	0 (0.19)	0.107 (0.16)	0.06* (0.031)	0.05 (0.035)	-0.035 (0.21)	0.078 (0.15)
Theft	0.069 (0.077)	0.063 (0.068)	0.104*** (0.014)	0.136*** (0.017)	0.05 (0.1)	0.043 (0.066)
Nonindex Crimes						
Drug	0.803*** (0.13)	0.876*** (0.12)	0.379*** (0.021)	0.533*** (0.026)	0.796*** (0.18)	0.721*** (0.11)
Simple Violent	0.03 (0.091)	0.065 (0.073)	0.136*** (0.014)	0.163*** (0.016)	0.035 (0.096)	0.067 (0.068)
Weapon	0.849* (0.43)	0.812*** (0.3)	0.479*** (0.053)	0.607*** (0.067)	0.712** (0.32)	0.707** (0.29)
All Other	0.183** (0.081)	0.163** (0.064)	0.126*** (0.012)	0.154*** (0.015)	0.192** (0.086)	0.147** (0.059)
Activities						
N. Street Stops	0.973*** (0.061)	1.101*** (0.057)	0.414*** (0.0098)	0.556*** (0.013)	1.051*** (0.082)	0.847*** (0.051)
N. Traffic Stops	0.754*** (0.062)	0.775*** (0.052)	0.396*** (0.01)	0.496*** (0.012)	0.693*** (0.067)	0.591*** (0.048)
N. Nonindex Arrests	0.645*** (0.084)	0.709*** (0.073)	0.322*** (0.013)	0.45*** (0.016)	0.584*** (0.092)	0.531*** (0.064)
N. Index Arrests	0.39** (0.2)	0.321* (0.17)	0.253*** (0.032)	0.321*** (0.039)	0.393 (0.24)	0.252 (0.17)
N. Uses of Force	0.631* (0.36)	0.646** (0.31)	0.484*** (0.057)	0.672*** (0.07)	0.605 (0.42)	0.648** (0.29)
N. Complaints	-0.407 (0.95)	0.608 (0.74)	0.362*** (0.12)	0.468*** (0.15)	-0.4 (0.99)	0.532 (0.65)
Costs ($)						
Activity Cost	0.547*** (0.097)	0.621*** (0.081)	0.332*** (0.015)	0.455*** (0.018)	0.514*** (0.11)	0.481*** (0.071)
Crime Cost (Full Murder)	-0.411 (0.56)	0.119 (0.41)	0.056 (0.077)	0.085 (0.092)	-0.426 (0.5)	0.196 (0.33)
Crime Cost (Low Murder)	-0.075 (0.15)	-0.061 (0.1)	0.124*** (0.018)	0.148*** (0.022)	-0.034 (0.12)	-0.021 (0.08)
Total Cost (Full Murder)	-0.337 (0.51)	0.158 (0.37)	0.077 (0.071)	0.113 (0.085)	-0.347 (0.46)	0.218 (0.3)
Total Cost (Low Murder)	0.009 (0.13)	0.032 (0.09)	0.152*** (0.016)	0.19*** (0.019)	0.038 (0.11)	0.049 (0.07)

Note: Table displays the implied elasticity estimates of the effect of the number of officers working in a district-hour on crime within that district-hour. Estimates are obtained by regressing outcomes on logged officer quantity (equations (12) and (13)), or using a control function ('CF') approach with a Poisson regression for either logs or levels of N. I exclude metrics of policing intensity and, in the 2SLS or CF approaches use the number of (or log of number of) predicted number of officers based on the rotation calendar prediction. All estimates are converted to elasticities depending on the specification. Levels OLS and 2SLS regression coefficients are divided by the mean of the dependent variable and multiplied by the mean of officer quantity; Log OLS and 2SLS divide estimates by the mean of the dependent; Poisson CF with Logs takes the estimate itself, and with Levels it multiplies it by the mean of officer quantity. Standard errors are clustered at the district-year-month-day-of-week-hour level and used to construct confidence intervals. Crime costs refer to total crime costs using Miller et al. (2021) incidence costs applied to reported crimes and 'Low Murder' refers to using a value of $400,000 rather than $8 million ('Full Murder'). Standard errors are clustered at the district-year-month-day-of-week-hour level and are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

C Additional Background and Data

C.1 Assignment and Attendance Data

The primary data for this study is daily assignment and attendance records for all units in the CPD between 2010 and 2019. 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.

Sample periods are occasionally used in this analysis. They are particularly useful to distinguish between pre- and post-scandal Chicago. In late 2015, video footage showing the murder Laquan McDonald by CPD officer Jason Van Dyke was released. This generated a massive scandal which also changed the oversight agency producing complaint data (see Rivera & Ba (2023)), reforms along with the ACLU lawsuit and new street stop policy causing a major change in street and traffic stops (Hausman & Kronick (2023)), among other reforms.

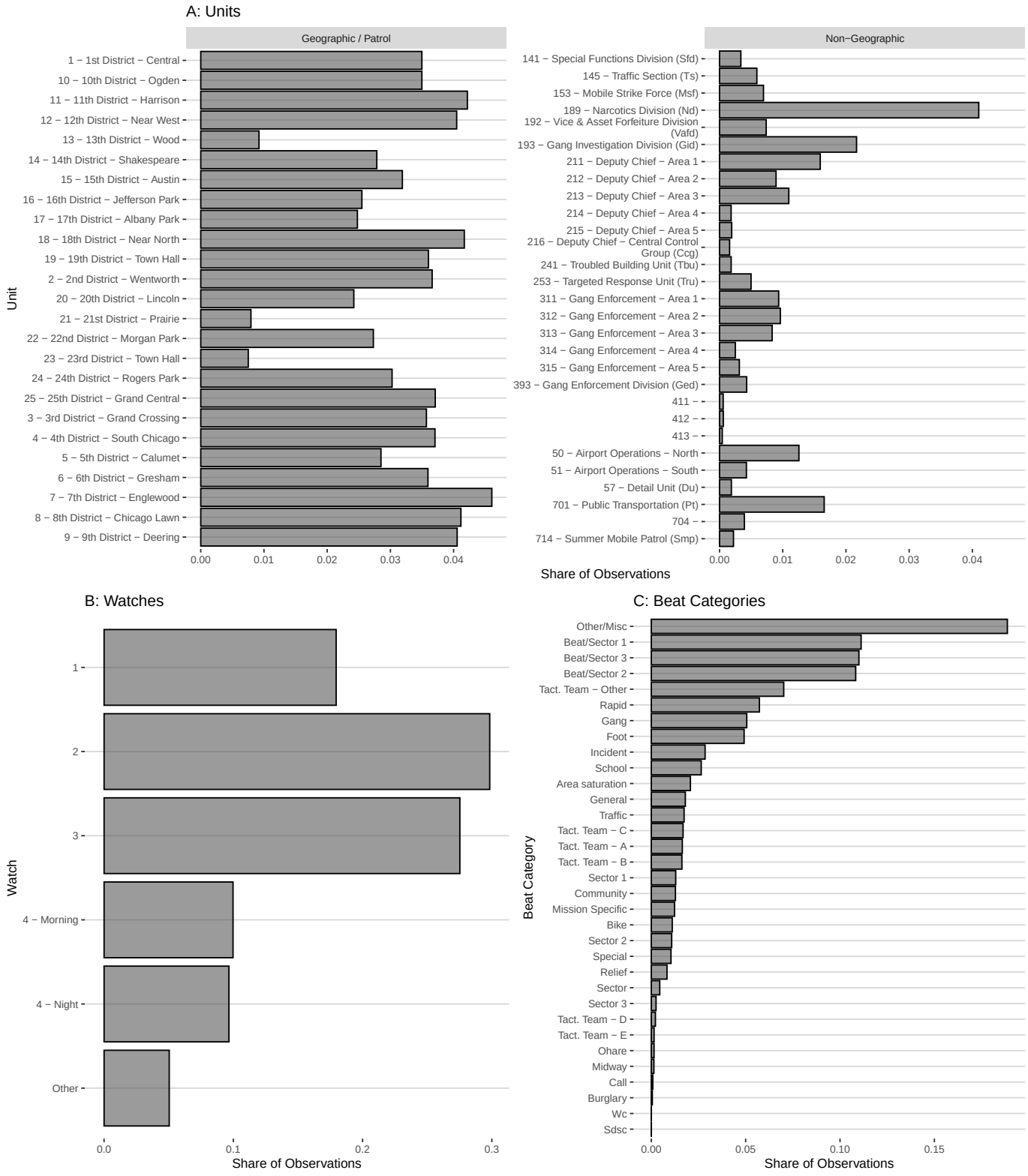
Initially, this data contains over 35 million entries including days-off. I restrict my sample to days where the officer is present for duty with a known assignment and to the sample to officers who are the lowest and most common rank of "police officer" (around 80% of the CPD). They must also be on 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).

The officer-shift analysis of 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 C.1 displays these units by name and number along with the number of officer-shift (day) observations per unit in the sample along with a break down of watches used and beat categories (classifications of exact beats to reduce the total number of unique beats).

As shown, most officers work in patrol/geographic units which map directly to districts. Police districts are large areas of the city (around 10 square miles each) similar to precincts in other cities. During this study, Chicago is divided into 22 (2013-2019) or 25 (2010-2012) 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.

Beats correspond to the officer's role which can vary widely. 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. Some assigned beats map directly to geographic beats, but this is not the norm. Instead, an assigned beat denotes both a general geographic area and an operational role.

Figure C.1: 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.

C.2 Officer Behavior and Performance

I link the daily assignment records to these officer activities based on shift start and end times and the date and time of each activity occurring and recorded officer(s). Panel C of Table 1 summarizes the means of the primary outcomes per 100 9-hour officer-shifts (900 officer-hours).

C.2.1 Street and Traffic Stops

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).

The data for both street and traffic stops start 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 late 2015, following a report on CPD stop and frisk by the ACLU, the CPD agreed to collect additional information on their stops replacing CC with ISR, which took effect on January 1st, 2016 and resulted in a significant decline in street stops.

'Traffic stops' are the second most common outcome. This data was obtained separately FOIA requests specifically for traffic stops rather than contact cards or ISR data, and the CPD often categorizes them separately. Traffic stops became significantly more common in 2016. 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 one officer per event.

C.2.2 Arrests

Arrest data include arrestee information, crime types, and arresting officers classified by 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 nonindex 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.

C.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 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). 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 2010 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 2010 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 to construct 2016-2019 complaint counts.

C.3 Mapping Activities to District-Hours

For stops and arrests, the raw data contain the district and day-hour in which the activity occurred. So one activity is assigned to the corresponding district-hour regardless of how many officers were involved in it in the officer-level data.

For uses of force, multiple officers can be involved and use force in a single incident, though the data also includes the district and time that it occurred at, though officers in the same incident may differ in the time their force use took place. I assign a single force incident at the district-hour level for all force uses reported to take place within the same district, hour, and RD number (which links incidents in CPD data).

For complaints, time data is more imprecise and occasionally clearly imputed, and in the 2016-2019 period district data is missing. I link complaints to times by determining which hour of the day overlaps with the officer's shift on that day. If multiple hours are possible, I assign the complaint to occur in a random hour that is within the set of possible hours. In 2016-2019, I impute districts as the unit of the officers on that day — so non-patrol unit officers' complaints are excluded.

C.4 Assigning Costs

C.4.1 Crime Costs

Crime costs are determined by Miller et al. (2021) incidence cost estimates and mapped to CPD crime types as shown in Table C.1.

Table C.1: 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.

C.4.2 Stop and Force Activity Costs

All activity cost estimates are conservative and likely underestimates. Prior work shows that stops incur costs from associated fines and tickets (Mello (2024)) and have negative effects on

education (Bacher-Hicks & Campa (2020a)). Use of force can generate significant spillovers in the forms of socioeconomic spillovers (Ang et al. (2021)), police scandals, police killings, downstream criminal justice costs, and costs to the department in the form of injuries which can result in increased officer aggression (Holz et al. (2023)). For example, a report from the LA Sheriff's Department's 2019 estimates each use of force costing $6,134 to the county alone (LASD (2019)).

C.4.3 Arrest Costs

I use values of $5,000 and $10,000 per nonindex and index arrest respectively, based on the Dobbie et al. (2018) estimates for the effect of detention relative to pretrial release and discounting the lower bound estimates significantly. A sizable portion of arrests in Cook County do result in detention, over 30% of the felony sample in Rivera (2023) are detained for at least 7 days with average detention costs of $22,000 due to very long duration in jail. Note that if only one out of every ten nonindex arrests results in a pretrial detention and pretrial release has no cost, the costs of the average nonindex arrest well exceed the estimated LM crime reductions. Also note that because of their infrequency, reducing index arrest costs to $0 does not significantly change total cost estimates.

However, this comparison also ignores the fixed costs associated with the arrest in the first place. Agan et al. (2023) finds that prosecution of non-violent misdemeanors can actually increase future criminal complaints and thus incur social costs. Even prosecuting a simple drug possession charge with the least costly outcomes and arraignment generates over $1,000 in case preparation costs alone (Rouhani et al. (2023)). Index arrests are more likely to result in detention and post-trial incarceration, which also yield large costs. A 6-month prison term can generate over $35,000 in social costs after taking into account incapacitation effects (Mueller-Smith (2015)), with a 1 year prison sentence costing $3,278 in lost earnings (Garin et al. (2023)) and longer sentences producing smaller marginal effects for crime-reduction (Rose & Shem-Tov (2021)).

C.4.4 Complaint Costs

I use data on settlements due to actions of Chicago police officers from 2010 to 2019 to provide a cost estimate for a complaint against the police. This does not include any direct harms or financial costs to the victims (Ba (2018)), time spent adjudicating the complaint, salaries of investigators, or downstream consequences of scandals such as lost trust and less reporting of crime (Ang et al. (2025)). 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 to almost $20 million, involving around 2,500 officers — totaling 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.

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.

Table C.2: Association between Officer Complaint Effects and Settlement Costs

	Settlement Cost per Shift	Settlement Cost
	(1)	(2)
Officer Complaint FE	12190.6***	5379333.8***
	(542.3)	(234860.1)
R2	0.002	0.0037
N. Obs	24274	24274

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. Column (1) uses total settlement cost (per named officer) by the number of shifts worked for the dependent variable and Column (2) uses 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. 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 C.2 displays the results. Based on Column (1), one additional complaint per shift is associated with $12,191 in settlement costs. This is consistent with regressing total settlement amounts (not per shift) on officer complaint effects, which produces estimates of $5.3 million per 1 unit increase in complaint effects per sample period (generally around 500 shifts).

C.5 Demographics Data

I link this assignment data to officer personnel records, containing demographic information (race, gender, age, tenure), as well as supplementary information on officers' experiences and skills, their education level and whether they speak Spanish. As shown in Table C.3, the

sample of officers is plurality white (48.1%), with the remainder being Hispanic (24.9%), Black (23.3%), and Asian/Native American/unknown ('Other race') (3.7%). As with many police departments, a majority of officers are male (80.9%). Officers tend to start in their late 20's around 28.9 years old, and the average officer in sample is around 41.7 years old and with 12.9 years of experience (tenure). Around 14.8% of the sample speak Spanish. For around 70% of the sample, I have educational history information, and 31% of officers have a BA or higher while 37.9% have less than a BA (e.g., associates degree, college credits).

Table C.3: Summary Statistics of Officer Demographics

	Pooled	2010 - 2012	2013 - 2015	2016 - 2019
	(1)	(2)	(3)	(4)
White	0.481 (0.5)	0.49 (0.5)	0.487 (0.5)	0.468 (0.499)
Black	0.233 (0.423)	0.252 (0.434)	0.24 (0.427)	0.209 (0.407)
Hispanic	0.249 (0.432)	0.223 (0.416)	0.237 (0.425)	0.282 (0.45)
Asian/Native American	0.0371 (0.189)	0.0347 (0.183)	0.0352 (0.184)	0.0409 (0.198)
Male	0.809 (0.393)	0.799 (0.401)	0.81 (0.392)	0.816 (0.387)
Female	0.191 (0.393)	0.201 (0.401)	0.19 (0.392)	0.184 (0.387)
Age	41.7 (8.25)	41.1 (8.02)	42.4 (7.83)	41.8 (8.76)
Tenure	12.9 (6.81)	12.1 (6.06)	13.5 (6.38)	13 (7.68)
Spanish Speaking	0.148 (0.355)	0.154 (0.361)	0.147 (0.354)	0.144 (0.351)
Edu >= BA	0.31 (0.462)	0.285 (0.451)	0.305 (0.461)	0.336 (0.472)
Edu < BA	0.379 (0.485)	0.343 (0.475)	0.378 (0.485)	0.414 (0.492)
Edu Missing	0.311 (0.463)	0.373 (0.483)	0.317 (0.465)	0.25 (0.433)

Note: Table presents summary of officer demographics and demographics in the assignment data.

C.6 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 15, 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).

Though the total number of officers working in a district-hour 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.

C.7 Crime

Reported crime is the 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. Table C.4 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 and Table C.5 displays counts and year-over-year changes in crimes, stops, and arrests.

Table C.4: 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	0.338	0.302	0.037	0.108
Nonindex Drug	0.123	0.000	0.122	0.997
Nonindex Simple Violent	0.302	0.237	0.065	0.217
Nonindex Weapon	0.020	0.005	0.015	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.

Table C.5: Crime, Stops, and Arrests by Year

	Crimes			Stops		Arrests	
	Violent	Property	Nonindex	Street	Traffic	Nonindex	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Counts							
2010	31001	122719	216673	265748	153950	121899	16605
2011	29658	121656	200556	297022	140829	113608	13036
2012	28484	115264	192377	380574	115646	107773	14030
2013	24628	102366	180278	464416	104076	106100	15171
2014	22659	86439	166421	481738	105072	96700	14077
2015	23229	81043	160153	503598	96802	85075	12916
2016	28277	87704	153405	72299	208656	62239	12598
2017	28064	89191	151367	79438	340865	61487	11730
2018	25883	87363	154892	105334	492575	64602	12004
2019	23952	81396	154342	120869	602962	70467	11780
% Change vs. Year - 1							
2010	-	-	-	-	-	-	-
2011	-4.3%	-0.9%	-7.4%	11.8%	-8.5%	-6.8%	-21.5%
2012	-4%	-5.3%	-4.1%	28.1%	-17.9%	-5.1%	7.6%
2013	-13.5%	-11.2%	-6.3%	22%	-10%	-1.6%	8.1%
2014	-8%	-15.6%	-7.7%	3.7%	1%	-8.9%	-7.2%
2015	2.5%	-6.2%	-3.8%	4.5%	-7.9%	-12%	-8.2%
2016	21.7%	8.2%	-4.2%	-85.6%	115.5%	-26.8%	-2.5%
2017	-0.8%	1.7%	-1.3%	9.9%	63.4%	-1.2%	-6.9%
2018	-7.8%	-2%	2.3%	32.6%	44.5%	5.1%	2.3%
2019	-7.5%	-6.8%	-0.4%	14.7%	22.4%	9.1%	-1.9%

Note: Table presents total number of crimes, stops, and arrests by year and the percent change relative to the prior year.