NBER WORKING PAPER SERIES

HOW FRONTLINE SUPERVISORS SHAPE PRIORITIES: EVIDENCE FROM POLICE LIEUTENANTS

Matthew Gudgeon
Andrew Jordan
Taeho Kim

Working Paper 35306
http://www.nber.org/papers/w35306

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

All opinions expressed in this manuscript are those of the authors. We thank Oeindrila Dube, Greg Distelhorst, Max Kapustin, Michael Makowsky, Derek Neal, Ben Pyle, Roman Rivera, Carly Will Sloan, Chris Stanton, Brittany Street, Anna Weber, and seminar participants at the NBER Crime Summer Institute, the Society of Labor Economists Annual Meeting, APPAM, SIOE, ASSA, Wharton People and Organizations conference, the University of Toronto, and West Point. All errors remain ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Frontline Supervisors Shape Priorities: Evidence from Police Lieutenants
Matthew Gudgeon, Andrew Jordan, and Taeho Kim
NBER Working Paper No. 35306
June 2026
JEL No. J15, J45, K42, M54

ABSTRACT

We study how frontline supervisors shape outcomes in public organizations with competing objectives and limited monitoring. We examine lieutenants in the Chicago Police Department, exploiting its rotational operations calendar for identification. We document dispersion in lieutenant fixed effects on team arrests and find race to be a key predictor. We then compare days Black and Hispanic lieutenants are predicted on duty to days supervised by white lieutenants. Teams under Black and Hispanic lieutenants make fewer low-level arrests while improving 911 call dispatch time. Lieutenants shape effort allocation by differentially granting overtime and awards for serious rather than discretionary arrests.

Matthew Gudgeon
Tufts University
Department of Economics
and NBER
matthew.gudgeon@tufts.edu

Andrew Jordan
Washington University in St. Louis
Department of Economics
awjordan@wustl.edu

Taeho Kim
University of Toronto
thk.kim@utoronto.ca

1 Introduction

Managers are important determinants of productivity and there is substantial heterogeneity in value-added across managers (Lazear et al., 2015). A growing body of empirical work estimates manager effects in private firms and explores what better managers do differently (Hoffman and Tadelis, 2021; Adhvaryu et al., 2023; Minni, 2026, for example). Recent work has also begun to empirically investigate the role of managers in the public sector, a setting in which hiring, firing, and pay incentives are often constrained (Rasul and Rogger, 2016; Janke et al., 2020; Fenizia, 2022; Best et al., 2023). While several of these find that manager influence extends to the public sector, others like Janke et al. (2020) do not. Furthermore, most existing evidence comes from settings with clearly defined objectives that can be well summarized by a single productivity metric (e.g., claims processed).

Much less is known about supervisory influence in environments characterized by multiple competing objectives with weak formal incentives and limited output measurement and monitoring. These conditions describe many high-stakes public-sector organizations, including healthcare, social services, and regulatory enforcement, where frontline workers exercise substantial discretion over multi-faceted and socially consequential decisions. Standard principal–agent models imply that supervisory influence could be limited in settings with multiple tasks and where monitoring and incentive tools are weak (Dixit, 2002).[1] It thus remains an open empirical question *whether* and, if so, *how* supervisors shape outcomes in such environments.

We study this question in the context of policing, where these organizational features are particularly stark. Police officers exercise enormous discretion over consequential decisions like whom to arrest. Police officers conduct multiple tasks with complex and potentially competing downstream impacts on outcomes like crime and public trust. Furthermore, supervisors in the district office cannot observe many officer actions. Formal incentives tend to be weak and often misaligned with broader organizational goals (Finan et al., 2017). We will show that, despite these challenges, mid-level supervisors — police lieutenants (LTs) — significantly shape team behavior and document specific ways they do so.

LTs are the highest-ranking front-line supervisors, and at least one must be on duty every shift. Watch operations LTs in particular oversee personnel and staffing, conduct roll call, review and approve all arrests, and manage day-to-day operations from district headquarters.[2] LTs operate under senior management from the district commander who oversees strategy but, unlike a watch operations LT, will not be present on every shift.

A large policing literature documents that individual officer characteristics shape policing behavior. For example, minority officers engage in less aggressive enforcement than their white peers (Ba et al., 2021;

[1]Indeed, Rasul and Rogger (2016) find that monitoring-intensive practices like performance-based incentives backfire for projects that require multitasking or when there are many principals.

[2]LTs can also serve as field LTs, responding to and assuming command of major incidents within their districts. Field LTs regularly assume the role of watch operations LT (such as on days off) and vice versa. LTs can also serve in a third role, that of tactical LTs. These LTs oversee tactical teams and do not commonly fill in for field or watch operations LTs. They work on a separate schedule that does not correspond to the normal three watches.

Goncalves and Mello, 2021; Hoekstra and Sloan, 2022; Rivera, 2022). At the organizational level, reforms like body-worn cameras, procedural justice training, and oversight agencies also shape outcomes (Adger et al., 2022; Dube et al., 2025; Mello et al., 2023). Much less is known, however, about the intermediate layer of supervision. A 2017 U.S. Department of Justice investigation into the Chicago Police Department, the focus of our study, emphasizes both the potential for and current deficiency of LT supervision (US Department of Justice, 2017). However, whether and how LTs shape outcomes remains largely unexplored empirically.

We begin with a descriptive exercise estimating watch operations LT fixed effects on daily, district-watch arrest outcomes, controlling for team (district-watch-year) and date fixed effects. We find meaningful dispersion of LT fixed effects even after adjusting for measurement error. Moving from the 25th to the 75th percentile LT is associated with a 0.18 standard deviation increase in daily team arrests. We then investigate what predicts this variation. Among a range of observable LT characteristics, including information on LTs' own enforcement and assignment histories as officers, race emerges as the strongest and most consistent predictor of LT effects. In particular, we find that Black and Hispanic LTs produce systematically fewer (0.06 standard deviations) low-level arrests than white LTs.

These descriptive patterns could reflect causal differences across LTs but may also reflect potentially endogenous sorting into shifts and exact days worked. To better isolate the causal effects of different LT features, we exploit a quasi-experiment generated by the CPD's predetermined rotation schedule.[3] The CPD's 22 districts follow an operations calendar that assigns all personnel—officers and LTs—to one of six day off groups (DOGs) determining who works each day. The schedule is set the prior November and goes into effect each January, generating exogenous variation in which LT is predicted to be on duty on any given day.[4]

We begin by re-estimating LT effects using only this schedule-induced variation and re-assess the extent to which they vary with observable LT characteristics. Specifically, we restrict attention to district-watch-year cells with exactly two predicted watch or field LTs. Using our day off-group-based predicted work schedules, we compare outcomes on days when one LT is predicted to work alone relative to days the other is predicted to work alone. Using only this within-team, predicted variation, we continue to find that LT race is a strong and meaningful predictor of differences in LT effects. Other covariates are generally not predictive, with the exception of an LT's prior public order arrest tendencies, which positively predict higher public order arrests (even after controlling for race).

The persistence of race as a key observable predictor of behavior is consistent with the large body of prior literature in this policing setting (e.g. Ba et al., 2021). Our findings suggest that these racial differences continue to matter even after officers are promoted to supervisory LT positions and are no longer in the field. Using the preceding results as motivation, we next turn to our primary empirical specification wherein we

[3]Recent, concurrent work similarly exploits this feature of the CPD, alone or in conjunction with other shocks, for identification (Ba et al., 2024; Ferrazares, 2025; Bollman et al., 2025).

[4]To ensure watch operations LT coverage, within a district-watch 2 LTs will typically be assigned to different, non-adjacent day off groups, such that for every 6 days, one works 2 days alone, 2 days together, and then has 2 days off while the other LT works.

isolate the causal effect of having a Black or Hispanic LT predicted to be on duty. We use race as a lens to understand supervisory influence because it is tractable to causal analysis, internally relevant and predictive, and directly connected to prior literature.

Before isolating the effects of having a Black or Hispanic LT on duty,[5] we first confirm that days predicted to have such LTs do not differ, on average, in terms of their crime conditions, predicted team composition, or the probability of having an out-of-rotation LT on duty. We show that predicted schedules are strongly predictive of actual presence. An LT's actual schedule could differ from his or her predicted schedule for several reasons including movement to a different shift, vacation, or sick leave. Nevertheless, we find that having a Black or Hispanic LT predicted on duty increases the probability of actually observing an LT on duty by 44 percentage points (from 14% to 58%).

On days when Black or Hispanic LTs are predicted to work, we find that subordinate officers make 0.13 (or 3.6% relative to the mean) fewer arrests per day.[6] This reduced form effect is substantial; it is equivalent to 6.5 times the reduction Ba et al. (2021) estimate when comparing Black police officers to white officers. The median shift in our data has 45 officers of which 6 typically are Black. As such, matching the effect of a minority LT would require nearly doubling the share of Black officers on a typical shift.

The decline in arrests is concentrated in lower-severity offenses. Index arrests — those associated with the most serious crimes — fall by a statistically insignificant 1.4%. In contrast, non-index arrests fall by 4%. Moreover, 88% of the total decline in arrests comes from non-index drug-related offenses (which fall by 6%) and public order offenses (-7%), which include disorderly conduct (-13%) and municipal code violations (-10%). The vast majority of arrest reductions are of Black persons, largely reflecting their share of baseline arrests. However, we are unable to reject differences in arrest rates (as a percent of the mean) across arrestee race. When we link to court outcomes, we also do not detect differences in conviction rates within arrest categories. Hence, within arrest categories, arrest declines appear to reflect typical as opposed to specific types of arrests.

These differences do not simply reflect reduced effort or officer activity. When we examine emergency dispatch times, we find that teams under Black and Hispanic LTs are dispatched faster to 911 calls, particularly high-priority calls where timeliness is most consequential (Vidal and Kirchmaier, 2018; Weisburd, 2021; Deangelo et al., 2023). We also find relatively tight null effects on the total number of field engagements recorded in dispatch-system data, together with evidence of reallocation away from social-order enforcement activities and toward community-assistance engagements, and no meaningful change in stop activity. These patterns reveal that minority LTs are, on average, making different operational tradeoffs among competing organizational objectives, placing less emphasis on low-level enforcement and more on emergency responsiveness. Supervisory influence in this environment operates through the reallocation of effort across competing objectives rather than solely through changes in total effort. Our results are consistent with a

[5]We pool Black and Hispanic LTs together because we generally cannot reject equality and pooling buys us power.

[6]An instrumental variables approach instrumenting for whether a Black or Hispanic LT is actually on duty yields an 8% reduction in arrests, though like Glover et al. (2017), we opt to focus on the reduced-form effect as the more conservative estimate, since the exclusion restriction that Black or Hispanic LTs only affect outcomes on days they are actually working is strong.

multitasking interpretation (Holmstrom and Milgrom, 1991; Dewatripont et al., 1999): when officers face multiple enforcement margins and a supervisor reweights the salience of measured outputs, agents reallocate effort across tasks without necessarily changing total effort.

Before turning to *how* supervisors shift officer effort across tasks, we mention several additional results that help characterize our main findings. First, effects are strongest for watch operations LTs — those stationed at headquarters who must approve arrests — and weakest for field LTs who are present in the field but do not have primary approval authority. This suggests that influence operates through something other than direct field supervision or modeling. Second, in officer-level specifications with officer fixed effects, we find that the same officer makes fewer arrests when working under a Black or Hispanic LT compared to working under a white LT in the same district-watch. This confirms that effects arise not because of different officers working under different LTs, but rather because officers adjust their behavior based on who is supervising on the given day. Third, we show that differences in other predicted LT characteristics (like gender, tenure, past assignments) generally do not affect outcomes. One exception is that LTs who as officers made more public order arrests tend to oversee teams that make more non-index and public order arrests. This is true even after conditioning on LT race, even though LT race remains the clearest observable predictor. Nevertheless, this shows that there are other classifiable LT 'types' beyond those captured by race.

How do supervisors manage to influence outcomes and officer decisions given the many constraints they face? In the last section of the paper we document several specific levers through which LTs signal their preferences. Watch operations LTs conduct roll call meetings at the start of the watch where they can communicate priorities; they oversee and approve all arrests; and they manage personnel including proposing awards and granting overtime. We show that Black and Hispanic LTs are more likely to grant valued overtime for extensions related to index arrests and less likely to do so for public order arrests. They are also less likely to nominate officers for commendations following lower-severity arrests. We find suggestive but imprecise evidence that they more often release arrestees without charge for low-level arrests, which is a rare outcome that in equilibrium need not be common (the threat alone could well be sufficient). Together, these findings suggest LTs rely on softer forms of influence — shaping expectations about what will be approved and rewarded — rather than direct monitoring. When organizational objectives compete and top management cannot specify clear priorities, frontline supervisors effectively operationalize tradeoffs through their approval and recognition decisions. The organization's actual priorities — as experienced by workers and observed in outcomes — are thus partly determined by frontline supervisors in charge of execution, not solely by leadership at the top.

The normative implications of weighting these competing objectives differently depend on the marginal costs and benefits of low-level enforcement (Owens, 2020), which could vary across place and time and depend on alternative support systems. While it is certainly possible that the marginal costs of pro-active low-level enforcement exceed the benefits, this need not be the case.[7] Our evidence cannot speak to the

[7]For example, on the marginal benefit side, Cho et al. (2024) and Agan et al. (2025) find no evidence of crime effects from reducing low-level arrests and charging rates for nonviolent misdemeanor offenses. Other work simultaneously suggests the marginal costs of low-level enforcement on the individual and on the community could be large and that minorities often bear the brunt of these costs (Agan et al., 2023; US Department of Justice, 2015, 2017). On the other hand, a lack of low-level enforcement could have

welfare implications of having one type of LT over the other. What we can establish is that frontline supervisors matter substantially in this high-discretion, weak-incentive setting, and that they exert influence through surprisingly soft mechanisms that operate via anticipated approval rather than direct observation or control.

We make three key contributions to the literature. First, we extend recent research on supervisory effects in private firms (Bertrand and Schoar, 2003; Bandiera et al., 2007; Lazear et al., 2015; Bandiera et al., 2020; Frederiksen et al., 2020; Hoffman and Tadelis, 2021; Adhvaryu et al., 2022, 2023; Metcalfe et al., 2023; Minni, 2026) to public sector settings characterized by extreme worker discretion, weak formal incentives, and multidimensional organizational objectives. We thus add to a small body of work on public sector managers by showing that supervisors can alter which of several competing objectives receives more weight (Rasul and Rogger, 2016; Janke et al., 2020; Fenizia, 2022; Best et al., 2023). Second, we fill an important gap in a large policing literature. Many papers have studied determinants of officer-level policing behavior, including a considerable body focused directly on the role of race (Donohue III, John and Levitt, 2001; Anwar and Fang, 2006; McCrary, 2007; Ba et al., 2021; Goncalves and Mello, 2021; Miller and Segal, 2012, 2019; Hoekstra and Sloan, 2022; Rivera, 2022). Other work has studied the role of executive officers in policing or broad organization-level policies.[8] Relatively less is known, however, about whether and how the middle layer of supervisors — who do not set organizational strategies, allocate resources, or influence hiring decisions — affect team tasks and outcomes.[9] Finally, we contribute to both the personnel and policing literatures by shedding light on how frontline supervisors shape worker behavior in high-discretion environments. While a growing literature documents substantial heterogeneity in manager effects, much less is known about the mechanisms through which managers influence outcomes, particularly in public-sector settings (Hoffman and Stanton, 2025). Beyond estimating supervisory effects, we observe concrete managerial decisions, including overtime allocation, commendation recommendations, and arrest oversight, that reveal how frontline supervisors shape officer effort allocation despite limited formal monitoring tools.

The rest of the paper proceeds as follows. Section 2 discusses the CPD, the role of LTs, and the day off group rotation calendar. Section 3 describes our data and sample. Section 4 presents motivating evidence on LT effects and their correlation with observable LT characteristics. Section 5 outlines our empirical strategy for identifying whether minority LTs affect team level outcomes and reports associated validity checks. Section 6 reports results. Section 7 examines which levers LTs use to exert their influence in this setting. Section 8 concludes.

long-run implications on even serious crimes (Chalfin et al., 2022).

[8] Kapustin et al. (2022) show that management quality of *executive* officers matters, documenting that district commanders in the CPD have broad influence over patrol behavior (e.g. stops) and respond causally to a management intervention designed to provide data-driven input into crime prevention strategies. Bulman (2019) uses time-series variation in the race of sheriffs to show that the ratio of Black-to-white arrests increases under white sheriffs. Other work shows that major interventions targeting all officers can affect behavior (Adger et al., 2022; Dube et al., 2025; Mello et al., 2023).

[9] There are some important exceptions: Smith (2025) studies the role of police sergeants, one level below the LTs we study, in the Dallas Police Department. Additionally, several other papers focus on specific aspects of police supervision: Rim et al. (2022) find that white police supervisors are less inclined to nominate Black officers for awards compared to white officers. Frake and Harmon (2023) find that officers working under supervisors with a high complaint history are more likely to receive complaints, and Bindler and Hjalmarsson (2023) show that female police supervisors in Seattle reduce subordinate gender gaps in frisking behavior.

2 Institutional Background: The Role of Lieutenants and the Day Off Group Rotation Schedule

2.1 CPD Organization and the Role of Lieutenants

Officers in the CPD work in one of 22 non-overlapping patrol districts (henceforth 'districts'), which span the geographical area of the city (see Figure A.1).[10] These districts perform core policing functions, and the majority of CPD officers (56% in 2019) work in one of these districts. In a typical patrol district, there are approximately 320 personnel. The CPD operates in three standard shifts called "watches" (nighttime, morning, and afternoon). Generally, the first watch runs from 10 PM to 6 AM the following day, the second watch runs from 6 AM to 2 PM, and the third watch covers 2 PM to 10 PM.[11] Our outcomes are constructed at the daily, district-by-watch (or shift) level (e.g. total arrests on a given day within that district-watch).

Frontline personnel include police officers and sergeants, who play essential roles in the day-to-day operations of law enforcement. At the lowest rank and serving on the front lines of law enforcement, police officers are responsible for various policing duties in the community. Their primary objectives include maintaining order, protecting life and property, responding to crimes and disturbances, conducting initial investigations, and apprehending suspects. Police officers constitute the majority of sworn officers, accounting for approximately 87% of personnel in CPD districts.[12] Sergeants serve as immediate supervisors who work closely with police officers to ensure adherence to all procedures and policies. They supervise, train, and coordinate the activities of officers, particularly at crime scenes, and serve as liaisons between upper management and officers in the field. Approximately 9% of district personnel hold the rank of sergeant.

This paper focuses on police LTs who are one rank above sergeants. 1–3 LTs manage district operations in *each* watch. LTs can serve in one of three roles (and sometimes switch between these roles): watch operations LT, field LT, or tactical LT. The crucial role of watch operations LT must be staffed at all times, so if one LT is working on a given shift they will be working as the watch operations LT. The watch operations LT oversees operations from the district's police station. He or she communicates key information and priorities at a roll call meeting that always occurs before each watch, oversees personnel/staffing, reviews and approves all arrest reports and complaints during the shift, and tends to other duties as outlined in the CPD directives (see Figure 1). Field LTs are tasked with responding to and assuming command of major field incidents within their respective districts, but will assume the role of watch operations LTs if that position is vacant (such as on the other LT's day off). Figure A.2 demonstrates that a significant number of LTs divide

[10] Before 2012, there were 25 districts, some of which merged in the early 2010s. Furthermore, we exclude district 5 because it operates under a special day off group arrangement. Hence our study sample focuses on 21 districts.

[11] Tactical officers (as opposed to beat officers) are assigned to what is sometimes called the fourth watch; in practice, this consists of a separate and distinct rotating schedule. Each district typically has three tactical teams, each consisting of ten officers and one sergeant. These three teams alternate between working (approximately) the second watch, the third watch, and providing relief for the other two tactical teams. The watch rotation schedule for each tactical team is specified annually.

[12] Excluding administrative officers on desk duty, there are two types of officers performing policing tasks in the street: beat officers are assigned to beats — small geographical areas comprised of a few blocks — and patrol these areas and answer relevant 911 calls. They make up around 80% of the officers with policing duties. The second category of officers comprises the tactical teams, whose patrol areas are not confined to a specific beat and vary each day depending on the prevailing crime conditions.

their workdays between serving as watch operations LTs and field LTs, whereas LTs assigned as tactical LTs — who oversee the officers in tactical teams — do not commonly assume the role of watch operations LTs. As a result of the overlap between field and watch LTs, many of the specifications in this paper will involve LTs who could be serving in either role, but we also employ approaches that focus more directly on the important role of watch operations LT.[13] The watch operations LT role effectively is effectively that of the highest level supervisor *that is always present* and can be thought of as the person in charge of ensuring that daily operations function smoothly.

Importantly, watch operations LT should be thought of as the highest-ranking front-line supervisor, as opposed to something like a senior manager. Each district is headed by a district commander (2 ranks above LT), who is in charge of strategic planning and crime suppression activities.[14] The commander is assisted by an executive officer — a captain (one rank above LT) — who is designated as second in command.[15] Since there is only one commander and one captain in each district, they cannot oversee the operations of all shifts which span the 24-hour day. Typically, commanders work Monday-Friday and captains Tuesday-Saturday, diverging from the regular day off group schedules followed by other ranks. As such, LTs operate under crime control strategies set by the ranks above them, but, as the top-ranking officer typically on duty, they retain wide discretion in how to go about their mission.

Studying the role of LTs (and watch operations LTs in particular) in district-watch level outcomes is of particular interest because they are the highest-ranked position that has to be filled on every watch — placing them in a key supervisory role with discretionary authority when it comes to overseeing personnel and arrests — but they simultaneously operate under higher levels of leadership and do not set formal policy. Whether and how this intermediate layer of supervision matters, particularly in a setting where the supervisor cannot directly observe most officer actions and police objectives can be complex and competing, is largely unexplored empirically.

2.2 CPD Operations Calendar and Day Off Groups

Our eventual identification strategy exploits variation in predicted LT work schedules (determined at the start of the year) that arises from features of the CPD operations calendar.

Because the CPD needs to operate every day, including weekends and holidays, it assigns officers (including watch operations and field LTs) to regularly scheduled day off groups. Most officers are assigned to one of six primary 8.5-hour day off groups (these groups are named 61, 62, ..., 66).[16] Officers in these groups

[13] We do not consider tactical LTs because it is challenging to incorporate tactical LTs' predicted schedules into our baseline empirical strategy. Tactical LTs do not adhere to the standard three-watch schedule; instead, they operate on a separate and less predictable schedule. As a result, Tactical LTs can report to work anytime of the day depending on the needs of the tactical teams. This stands in contrast with their tactical team subordinate officers who work rotating watches every police period.

[14] Specifically, the district commander is "accountable for the performance of all district personnel and the actions taken to address police service and crime suppression activities within a district" (Chicago Police Department Special Order S03-03: District operations).

[15] The executive officer "assists the district commander in areas of operation, administration, strategic planning, and evaluation." (Chicago Police Department Special Order S03-03: District operations)

[16] Tactical team officers are also assigned the 60s series day off groups. Other special officers can be assigned to different day off group series such as the 70s or the 1-7 series.

adhere to a pre-determined rotation: they work for four consecutive days followed by two days off, cycling through this pattern every six days. The day off groups are structured such that, on any given day, two of them are off-duty while the remaining four are on active duty.

Figure A.3 presents the operations calendar for the year 2015 for illustration purposes. The designated day off groups on break are marked in the lower left-hand corner of each date square. For example, an officer belonging to the 65 group would have days off on Tuesday January 6th and Wednesday January 7th. In the subsequent week, they would be off on Monday January 12th and Tuesday January 13th, and so on, cycling through this pattern, concluding with days off on Saturday, December 26th and Sunday, December 27th. LTs in any other day off group will have different scheduled days off.

Additionally, the calendar is set up so that two officers in the same day off group will work all of their 4 working days out of every 6 day cycle together, officers in adjacent day off groups (e.g. 62 with 61 or 63) will work 3 out of every 6 days together, and officers in non-adjacent day off groups (e.g. 62 and 64) will work 2 out of every 6 days together. This means that each officer will work with all other officers regardless of day off group, but they will work with officers in some day off groups more frequently than others. Since watch and field LTs are also slotted into one of these 6 day off groups, this means that watch operations LTs will work with all officers assigned to that shift at some point in time, but they will work with some more frequently than others. Moreover, there are typically 2 field/watch operations LTs in a given shift, and they will typically be assigned non-adjacent day off groups. Thus, out of every 6 days, each LT typically works 2 days alone (as watch operations LT), works 2 days with the other LT (where one would be a watch LT and the other a field LT), and takes 2 days off.

How are officers assigned to day off groups in a given year? In November of the previous year, police officers submit their preferences for day off groups (which they will begin adhering to in January). The district commanding officers distribute officers across day off groups, respecting officers' seniority and preferences, but most importantly, making sure scheduled manpower is balanced across days. This means that officers and LTs could potentially influence their day off group. For example, LTs could theoretically choose their day off group so as to avoid working on a given day (perhaps an anniversary or a key holiday). Crucially, however, they cannot choose their schedule so as to avoid working on *all* such days. For example, day off group 63 in 2015 would not be scheduled to work Christmas, but would have been scheduled to work New Year's. Similarly, a day off group that has more weekends off in a given month will have fewer weekends off in other months. Likewise, while you could perhaps choose a schedule that avoids one high expected crime day, it is not possible to do so for all. We will argue below (and support empirically) that due to the 4-day/2-day rotation, the days on which a Black or Hispanic LT is predicted to be working based on the CPD operations calendar do not, on average, have different expected crime rates than the days they are scheduled to be off.

Constructing and Validating Predicted Work Schedules In order to construct predicted work schedules, we assign each officer and lieutenant a day off group for the year based on January records. Specifically, we assign them their day off group that was in effect at the end of January, based on a day off group history file obtained through the Freedom of Information Act (FOIA) that records each officer's day off group and the effective date of each change. We use this day off group, together with the operations calendar, to predict whether each officer or LT was scheduled to work (i.e. not have a day off) on a given day during the remainder of the year. We further assign officers and LTs to a specific district-watch for the year using their modal district-watch assignment in January, as observed in daily attendance records. This allows us to predict whether a given district-watch-day has a specific officer or LT on duty for the remainder of the year. Throughout the paper, we focus on daily district-watch level outcomes from February through December of each year using our January-based predictions to avoid potentially more endogenous mid-year changes.

Officers' and LTs' actual work schedules will vary from their predicted schedules for several reasons. First, officers might be reassigned to different watches or districts, change roles, or exit the force. Second, it is possible for officers to change their day off groups over the course of the year. Indeed, one of the roles of the watch operations LT is to ensure day off group balance. There are a number of reasons officers can be pulled out of work (e.g. leaving the district, medical leave) causing potential imbalances in day off groups that may need to be remedied by switching day off groups around.[17] Third, officers could work on a scheduled day off or — and more likely — not work on a scheduled work day. Days off act like weekends. In addition to days off, officers and LTs will take leave and be absent for other duties. Hence, even when we predict them to be on duty (not have a day off), there are many reasons why they may not actually be working that day.

Our day off group-based predictions nevertheless perform well. The main sources of error in practice are absences due to leave and reassignments to different shifts. We assess prediction accuracy using actual attendance and absence records, which include a specific code for "DAY OFF" that is distinct from other leave types.

For each day between February and December that we observe an LT in the CPD data, we construct two indicators: whether they had a predicted day off, and whether they actually took a "DAY OFF" according to absence records. When we restrict attention to days on which LTs remain in the same district-watch as in January, the two indicators agree 93% of the time. This confirms that day off groups are highly stable and reliably predict scheduled days off. Without restricting to LTs who stayed in the district-watch, we correctly classify 84% of days off.[18] However, some of these correct classifications come from LTs who changed shifts. If we further count all days in which the LT is no longer in the same district-watch as incorrect predictions, the overall accuracy falls to 67%. In short, day off groups are highly predictive within a shift, but movement across shifts reduces predictive power.

Furthermore, correctly predicting scheduled days off is not the same as predicting whether an LT is actually present for duty — officers can also be absent on non-day-off days due to leave or other reasons. When we shift the goal from predicting scheduled days off to predicting actual duty presence, accuracy falls across all comparisons. Restricting to LTs who remain in the same district-watch, we correctly predict duty presence

[17] In such cases, the commanding officers ask for volunteers to change their day off groups, and if there are not enough volunteers, they would reallocate officers based on seniority.

[18] On average, LTs spend about 72% of observed days in the same district-watch as identified in January.

79% of the time, compared to 93% when predicting scheduled days off alone. Without any district-watch restriction, accuracy drops to 70% for duty presence versus 84% for days off. Finally, when we count all days in which the LT is no longer in the same district-watch as incorrect predictions, accuracy falls to 57%, compared to 67% when predicting days off.

Overall, our predictions are meaningful but imperfect, with the two primary sources of error being officers moving to other shifts and officers taking additional leave beyond their scheduled days off.

3 Data and Sample

Data. We use personnel data from the CPD, obtained by submitting numerous FOIA requests. We focus on the period from 2013-2020, a timeline chosen due to significant organizational changes in the CPD around 2012.[19]

Our data include demographics, ranks, and the unit assignments of all sworn officers in the CPD from 1946 to 2020. Between 2010 and 2020, we also have daily attendance and assignment data that allows us to observe unit, hours worked, and the reason for any absence (e.g. scheduled day off or personal leave). Additionally, we obtained historical day off group calendars for each year from 1972 to 2020, which tell us which days a given day off group is supposed to be working, and day off group histories at the officer level, which include the effective date of change.

Our primary arrest outcome data come from case-level arrest logs. Between 1999 and 2020, for each arrest incident, we observe the arresting officer(s), charges, FBI code for the crime, demographics of the subjects, incident time, and location. Based on the FBI code, we can categorize each arrest as index and non-index. This classification approach, commonly adopted by US police departments including the CPD, is consistent with the literature (e.g. Mello, 2019; Rivera, 2022). Index arrests encompass more serious crimes: homicides, sexual assault, robbery, aggravated assault and battery, burglary, larceny, motor vehicle theft, and arson. Non-index arrests include all the other arrests and are less severe. We also construct a public order arrest category, as in Premkumar (2021) and Kim (2022), which is a subset of non-index arrests comprising typically low-level offenses. Specifically, it covers arrests relating to liquor laws, disorderly conduct, miscellaneous non-index offenses, municipal code violations, and traffic violations. Finally, we use Dube et al. (2025)'s pre-registered 'discretionary' arrest grouping, which is intended to capture arrests that are "discretionary and plausibly unnecessary." It includes "charges such as obstructing an officer, resisting an officer, disobeying an officer and various types of disorderly conduct" that may "occur in situations where the officer may be responding emotionally, out of irritation or frustration (for example, based on their perception that a subject is being disobedient or disrespectful)" (Dube et al., 2025).

We supplement these sources with additional administrative datasets on policing activity and operational

[19] In 2011-2012, the CPD restructured its operations, reducing the number of police districts from 25 to 22. Additionally, prior to 2012, the role of watch operations LT was known as watch commander, a position that could be occupied by both captains and LTs. Post-restructuring, the CPD differentiated the roles of captains and LTs, assigning captains as executive officers and LTs with the direct management of district operations.

outcomes. These include dispatch data from the Chicago Office of Emergency Management and Communications (OEMC) Police Computer Aided Dispatch (PCAD) system (2013–2020), as well as data on use-of-force incidents, complaints, stops, 911 calls, and crime incidents (2013–2020). The OEMC dispatch data record police event assignments generated through both 911 calls and officer-initiated activities. For each event, we observe the event category, dispatch timestamps, responding units, and incident location information. We use these data to construct measures of dispatch times and classify incidents by call type. Use-of-force incidents come from Tactical Response Reports, which officers are required to complete whenever force is used. For investigatory stops, we observe the date of the stop, the responsible officer, demographic descriptors of stopped individuals, and the time and location of the incident.

We also merge with 1984-2019 court docket data from the Cook County Circuit Court, which has jurisdiction over the entirety of CPD's patrol area (Jordan et al., 2023). We link court records to arrests using central booking numbers. When an arrest does not appear in the court records, we mark it as not being taken up by the prosecutor. We then use the disposition of the matched court cases to identify those that ended in a formal conviction rather than dismissal by the prosecutor or judge or acquittal in court. When studying court outcomes, we restrict our sample to the years 2013-2016 (instead of 2013-2020) because we need follow-up time to observe conviction.

In Section 7, when we assess how LTs influence outcomes, we utilize additional datasets covering release without charge decisions, overtime approvals, and awards received. We discuss these in more detail in that section.

Our primary unit of analysis is the district-watch-day level. Hence, we aggregate outcomes like total daily arrests within a unit and watch (using the relevant timestamps). Since we reserve January for assigning each officer their day off group, our analysis data covers daily district-watch level outcomes between February and December of each year from 2013 to 2020.

Summary Statistics Table 1 presents summary statistics for our analysis sample. The dataset comprises 168,462 observations spanning the years 2013 to 2020 and the days from February 1 to December 31. Each observation represents a daily, district-by-watch level outcome. The data encompasses 63 district-watch combinations covering 504 district-watch-years.[20] On average, 1.3 LTs are on duty during any given period, and 20% of the observations are supervised by Black or Hispanic LTs. The typical district-watch has approximately 3.5 daily arrests, with an average of around 0.5 index arrests and 3 non-index arrests per day.

Accounting for all officers who served as LTs between 2013 and 2020, our sample includes 442 LTs who worked as either watch operations or field LTs across district-watch combinations. Of these, 122 are Black or Hispanic. Table A.1 compares Black and Hispanic LTs to their white counterparts across a range of characteristics, including gender, tenure, awards, complaints, and prior performance measures as officers in earlier ranks. See Section B.1 for details on variable construction. We find relatively few significant

[20]Our analysis sample includes 21 districts and 3 watches. We exclude district 5 because it operates under a special day off group arrangement.

differences across these dimensions.[21]

4 The Role of Lieutenant Supervisors

Before turning to our primary identification strategy that estimates causal differences in team outcomes across LT race, we first present motivating evidence that LT supervisors matter for team arrest outcomes. We begin by estimating LT fixed effects directly, which we interpret as descriptive summaries of persistent differences in team arrest behavior across supervisors. We find meaningful heterogeneity across LTs. We next show that some of this heterogeneity is systematic and can be predicted by observable LT characteristics, with LT race in particular emerging as a key predictor. The preceding patterns could reflect causal effects but could also be confounded by selective sorting into assignments. In Section 4.3, we show that these findings persist when using an approach that mirrors our eventual identification strategy and only leverages within-team variation in LTs' predicted work schedules for identification. Taken together, these results support a key role for LTs and motivate our focus on race as a lens for understanding supervisory influence in the remainder of the paper.

4.1 Dispersion in Lieutenant Fixed Effects

We begin by estimating LT fixed effects among watch operations LTs based on their actual attendance records. These records tell us who is actually serving as watch operations LT on any given shift, allowing us to construct an indicator for all the shifts that any given LT served as a watch operations LT. We estimate the following specification:

$$Y_{jt} = \alpha_{l(jt)} + \gamma_j + \delta_t + \varepsilon_{jt}, \tag{1}$$

where Y_{jt} is the daily arrest count for team j on date t, $\alpha_{l(jt)}$ is a fixed effect for the watch operations LT l supervising team j on date t, γ_j is a team (district-watch-year) fixed effect, δ_t is a calendar date fixed effect, and ε_{jt} captures uncorrelated noise in the outcome. The LT fixed effect $\alpha_{l(jt)}$ captures the residual component of arrest behavior associated with a given supervisor after absorbing permanent differences across teams and common time shocks. We treat these estimates as descriptive — they are estimated based on both variation in actual duty presence within a district-watch-year and variation based on moves across shifts. Formal identification would require that LT assignment and duty presence be uncorrelated with unobserved determinants of arrests conditional on team and date effects. We will relax this assumption in Section 4.3.

Because estimated fixed effects incorporate sampling noise and mechanically overstate true dispersion, we apply an empirical Bayes shrinkage procedure.[22] The left panels of Figure 2 display the resulting, shrunk

[21] In this table, we exclude 9 Asian/Pacific Islander (API) LTs.

[22] We follow Weisburst (2022) and decompose the variance of estimated LT effects into a signal component $\hat{\sigma}^2_\alpha$ and a noise component $\hat{\sigma}^2_\varepsilon$ via the moment-based correction $\hat{\sigma}^2_\alpha = \hat{\sigma}^2_v - \hat{\sigma}^2_\varepsilon$, where $\hat{\sigma}^2_v$ is the total variance of the estimated effects. The shrinkage-adjusted effect is then

$$\tilde{\alpha}_l = \left(\frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \hat{\sigma}^2_{\varepsilon,l}/N_l} \right) \hat{\alpha}_l,$$

distributions, standardized by the standard deviation of each outcome, for watch operations LTs with at least 100 days of supervision. Two patterns emerge:

First, there is substantial heterogeneity across LTs: the interquartile range of effects for all arrests spans approximately 0.18 standard deviations of the outcome. This corresponds to 0.66 daily arrests, or 19% of the mean.

Second, dispersion is not uniform across arrest outcomes. The distributions are visibly wider for lower-level and more discretionary enforcement outcomes — public order and non-index arrests — than for serious index crimes. Noise-corrected signal variances ($\hat{\sigma}^2_\alpha$) reported in Appendix Table A.2 confirm this pattern: the normalized signal standard deviation is 0.20 for non-index arrests compared to 0.09 for serious arrests. This is consistent with supervisory influence operating most strongly in areas where officers retain the greatest individual discretion and organizational objectives are least tightly specified. By contrast, index arrests are more tightly constrained by underlying criminal incidents, investigative protocols, and evidentiary requirements, leaving less scope for supervisory preferences to shift outcomes at the margin.

4.2 Systematic Correlates of Lieutenant Effects

We next ask to what extent these LT effects are correlated with observable LT characteristics. The right panels of Figure 2 report coefficients from regressing the shrinkage-adjusted LT fixed effects on observable LT characteristics, weighting by total days supervised. For ease of comparison across characteristics, we convert continuous characteristics (e.g. time to promotion) into indicator variables equal to one if the characteristic is above the sample median and report standardized effects. We also summarize the (unstandardized) results in Table A.3, which reports results from a series of bivariate regressions, and Table A.4, where we regress LT effects on the covariates simultaneously.

We find that LTs who are Black or Hispanic are clearly and systematically associated with being a low-arrest LT. This link is strongest for more discretionary outcomes — non-index and public-order arrests — and is absent for serious index arrests, consistent with the dispersion patterns documented above. These effects are also large: Black and Hispanic watch operations LTs supervise teams that make around 0.06 standard deviation fewer arrests, or 0.23 fewer arrests per day (6.5% of the mean). This is comparable to about a third of the effect of moving from the 75th to the 25th percentile of watch operations LTs.

The fact that LT race is systematically and strongly correlated with cross-LT variance in enforcement behavior is consistent with a large literature documenting systematic racial differences in officer behavior at lower ranks, with minority officers generally making fewer arrests and engaging in less discretionary enforcement (e.g., Ba et al., 2021). Our results suggest that the bundle of characteristics associated with race continues to be relevant once the officer rises up in the ranks and leaves the field for the district office.

Other observable characteristics generally display weaker and less precisely estimated relationships with the LT fixed effects. The most consistent other predictor is also related to race: LTs who work in units where

where N_l is the number of team-days supervised by LT l. Intuitively, LTs whose effects are estimated more precisely receive less shrinkage, while LTs observed on relatively few team-days are shrunk more heavily toward the overall mean.

their demographic background more closely matches the neighborhoods they police make fewer arrests.[23] LTs with prior detective experience also tend to exhibit lower fixed effects for non-index and public-order arrests. In the CPD, as in many other departments, becoming a detective is a competitive position that requires passing a promotional exam, even though it does not entail a change in pay grade, and hence might hint that LTs of higher 'quality' also make fewer arrests. Unlike the other covariates we consider, which have complete or near-complete coverage, we only observe LTs' prior arrest histories for 65% of LTs who were still serving as officers when our data begins.[24] This could make it harder to detect significant differences in LT effects across LTs' prior enforcement records. In Tables A.3 and A.4, we see that an LT's prior arrest history appears positively correlated with Dube et al. (2025)'s discretionary arrest category but not significantly with other arrest categories.

As is common in settings with limited observable supervisor characteristics, the observables jointly explain a modest share of the LT effects themselves (Bertrand and Schoar, 2003): the R^2 is 6.3% (or an adjusted R^2 of 3.6%). However, simply dropping LT race and racial match with the district lowers these to 3.5% (1.3%). Thus, LT race stands out both in terms of its relatively large and economically meaningful effect size and its standalone predictive power relative to other observable LT characteristics (including past performance as an officer).

4.3 Moving to Causal Identification

The preceding fixed-effects variation and its correlates could reflect either causal supervisory effects or endogenous sorting of certain types of LTs into specific teams and work days . To assuage selection concerns, we next take an approach that mirrors our eventual, primary identification strategy and leverages only within-team variation in predicted LT schedules.

Specifically, we restrict attention to the 265 district-watch-year cells with exactly two predicted watch or field LTs based on LTs' January assignments. Using our day off-group-based predicted work schedules, within each cell we identify days when one LT (LT1) is predicted to work alone (and hence as a watch operations LT), days when the other LT (LT2) is predicted to work alone, and days when both are predicted to work simultaneously. We compare outcomes on LT1-alone days relative to LT2-alone days, separately controlling for the both-present days. We focus on LT1 versus LT2 solo days to better isolate the role of watch operations LTs. When using predicted schedules we cannot be sure whether the LT actually serves as a field or watch operations LT, but if they end up on duty alone they will serve as a watch operations LT. Specifically, we estimate:

$$Y_{jt} = \alpha_j \, \mathit{LT1alone}_{jt} + \beta_j \, \mathit{LTstogether}_{jt} + \gamma_j + \delta_t + \varepsilon_{jt}, \tag{2}$$

where Y_{jt} is the daily arrest count for team (district-watch-year) $j \in \{1, \ldots, 265\}$ on date t; $\mathit{LT1alone}_{jt}$ equals one on days when LT1 is predicted to supervise team j alone and zero otherwise; $\mathit{LTstogether}_{jt}$

[23] To measure racial matching, we define whether an LT belongs to the largest racial or ethnic group in the district they supervise and then average this measure across assignments for each LT.
[24] See Section B.1 for additional details on variable construction.

equals one on days when both LT1 and LT2 are predicted to be present in team j and zero otherwise; and the omitted category is days when LT2 is predicted to be alone in team j. The coefficients α_j and β_j are team-specific, yielding 265 distinct estimates of each. As before, γ_j is a team (district-watch-year) fixed effect, and δ_t is a calendar date fixed effect. This produces 265 estimated relative fixed effects for one LT compared to the other within the same team environment.[25]

This specification minimizes identification concerns such as endogenous sorting into particular teams, neighborhoods, or workdays. It is identified as long as LTs' predicted schedules are uncorrelated with other determinants of daily district-watch-level outcomes. We find support for this assumption in the context of our eventual reduced form approach in Section 5. At the same time, by using only predicted as opposed to actual LT variation, this specification could underestimate effects.

In Table A.5, we show the one-by-one correlation between the estimated shrunken, relative LT FEs and the difference in characteristics for those same LTs. In Table A.6, we again include these covariates simultaneously as opposed to one at a time. Mirroring the earlier more descriptive approach, LT race continues to be a consistent, significant, and strong predictor of LT effects. Black and Hispanic LTs supervise shifts during which between 0.12 and 0.16 fewer daily arrests are made (3.6-4.8% of the mean). These effect sizes remain substantial, though they are smaller than 0.22, likely in part because this specification leverages predicted duty presence rather than actual presence. The correlation between LT race and arrest FEs is absent for serious index arrests and concentrated in non-index and, especially, public order arrests. The persistence of the correlation between LT race and LT effects in this design suggests that the relationships documented above do not simply reflect endogenous assignment of LTs to different types of teams or neighborhoods.

The other covariates are generally not consistent predictors of LT effects. Racial match is no longer statistically significant when using this within-team empirical variation. Prior LT public order arrest history is an exception. It is now positively correlated with being an LT that supervises teams that also end up making more public order arrests. This holds even conditional on differences in LT race, suggesting that at least some of the *within-race* differences across LTs can be captured using observables based on prior LT records. Together, the covariates have an adjusted R^2 for arrests of 7.8%.

Taken together, the preceding approaches motivate our focus in what follows. In the remainder of the paper we estimate the causal effect of assignment to a Black or Hispanic LT on team-level outcomes. We use LT race as a lens to identify the component of supervisory variation that is tractable to causal analysis, highly relevant in our sample, and directly connected to the broader literature linking policing patterns to officer race and ethnicity.

[25]The LT1 versus LT2 labeling within each team is arbitrary; we assign it by taking the first LT id number when sorted by team and day. This labeling is immaterial for the second-stage analysis where we regress each relative fixed effect on the same-ordered difference in LT characteristics, so flipping the labels for a given team flips the sign of both the outcome and the regressor.

5 Empirical Strategy

5.1 Estimating The Reduced Form Effect of A Predicted Black or Hispanic LT

Motivated by the preceding findings, we now present an empirical strategy for identifying the causal effect of having a Black or Hispanic LT predicted on duty and report validity checks in support of our identifying assumptions. Specifically, we compare daily, shift-level outcomes (e.g. total arrests) on days a Black or Hispanic LT is predicted to be at work to days in which all Black or Hispanic LTs are predicted to be off-duty.

We rely on predicted schedules derived from day off groups to address potential endogeneity in when LTs actually work. There are many reasons why LTs may work on days other than those we predict — LTs can take leave, move to a different district, watch, or day off group, or get promoted — and these actions could at least in principle be associated with both team performance and potentially race. While the police setting has several features that might minimize such concerns, we note that these concerns could be more salient at higher ranks, since lower ranks may find it harder to get approved leave.[26]

Formally, we estimate the following reduced form specification for district-by-watch-by-year j on day t.

$$\begin{aligned} y_{jt} = {} & \beta_1 \text{Any Predicted Black/Hisp. LT}_{jt} \\ & + \beta_2 \text{Any Predicted Male LT}_{jt} + \beta_3 \text{Predicted Num of LTs}_{jt} + \alpha_j + \delta_t + \epsilon_{jt}. \end{aligned} \tag{3}$$

Our outcomes of interest, y_{jt}, are measured daily at the district-watch level, such as total arrests. *Any Predicted Black/Hisp. LT* is a dummy variable for whether any Black or Hispanic watch operations or field LT is predicted to be working based on the operations calendar.[27] α_j denotes district$\times$watch$\times$year fixed effects. This ensures that all daily comparisons are made *within* a district-by-watch-by-year. As in Section 4, we also include separate date fixed effects (δ_t). These help absorb common shocks to underlying crime rates. Throughout the paper, we control for whether any male watch operations or field LT is predicted to be working (since race and gender of LTs are correlated), and for the number of watch operations or field LTs we predict to be working (i.e. the number of LTs that show up in the district-watch in January's attendance records) to allow for potential differences in team outcomes when two LTs are working as opposed to just

[26] An anecdote from Baltimore reinforces the idea that LTs might have more discretion to deviate from their schedules: "New Year's Eve is a hair-raising night to work. Traditionally many residents celebrate by shooting guns in the air. ... Commissioner Edward Norris, in his first year in Baltimore, decided to police New Year's Eve aggressively. He canceled all regular and vacation days off. The shift commanding LT called in sick. He said he had hives. ... The following roll call, on New Year's Day, the LT returned in high spirits and with no sign of hives" (Moskos, 2008).

[27] While there is no variation in *Any Predicted Black/Hisp. LT* in teams without a Black or Hispanic LT or teams with only Black or Hispanic LTs, these teams contribute towards estimating the coefficients on control variables. In practice, restricting to teams with variation in *Any Predicted Black/Hisp. LT* changes the coefficients and standard errors little, as we show later.

one.[28] We cluster standard errors at the district-watch-year level.[29]

Our focus is on LT race, but the effect of LT gender (and the associated coefficient on Any Predicted Male LT) is certainly of independent interest. In general, we do not detect statistically significant differences across gender. We report these effects in Section 6.3, where we also probe how outcomes vary with other, observable LT characteristics besides gender and race.

If the unobservable determinants of daily district-watch level outcomes are uncorrelated with having a predicted Black or Hispanic LT conditional on controls, β_1 can be interpreted as the causal effect of having a Black or Hispanic LT *scheduled to be on-duty* relative to district-watch level outcomes on the days all such LTs are scheduled off duty.

Importantly, this strategy identifies reduced form effects that are both contemporaneous and relative. We elaborate on each point in turn. First, this design recovers the direct, contemporaneous effect of a predicted Black or Hispanic LT on a particular day, contrasting how officers behave when that particular LT is predicted on duty to how a similar set (or, in a later officer level specification, the same set) of officers in the same district-watch behave on days under a different predicted LT. These same-day changes in behavior could reflect, for example, the expectation and standard setting that occurs at the start of the watch during the roll call meeting. On the other hand, this approach will underestimate effects that persist beyond a given shift. If Black or Hispanic LTs influence officers in ways that carry over to days when they are predicted absent, we will not fully capture that. We present some heterogeneity results suggesting effects may linger, but identifying long-run impacts of minority LT exposure would require a different identification strategy and is left for future research.

Second, we prioritize a reduced form approach over an instrumental variables (IV) approach, though we estimate both and discuss the relevant first stages as important context. The IV approach requires an exclusion restriction that Black or Hispanic LTs affect team outcomes only on days they are actually present. This restriction could well be violated in our setting. Because the day off rotation schedule means some officers work more frequently with certain LTs, repeated interactions may lead officers to carry those LTs' expectations and norms into shifts led by a different, one-off substitute LT. Since we expect at least some persistence in LT-instilled expectations, culture, or norms, we follow Glover et al. (2017) and opt for the more conservative reduced form approach. This is also the most policy-relevant estimate for the thought experiment of assigning a new watch operations LT to a district, given the rotation-based nature of police scheduling.

Finally, all effects are estimated relative to not having a predicted Black or Hispanic LT on duty. Because the watch operations LT position needs to be staffed, this position is typically filled by any watch or field LT who is on duty. If no LT is available, a substitute LT from another district or watch could fill in. In

[28] We control for the number of LTs linearly as opposed to as a fixed effect, but in practice this makes no difference whatsoever, since among teams that provide our identifying variation (i.e. for which *Any Predicted Black/Hisp. LT* switches off and on over the course of the year), almost all have 2 predicted LTs.

[29] The day off groups effectively assign LTs to work on close to random sets of days within each district-watch-year. Hence, we also report randomization inference p-values based on randomly reassigning LT day off groups. These yield p-values close to those obtained when clustering at the district-watch-year.

practice, in the majority of cases, the relevant counterfactual involves having one or more white LTs actually on duty.[30] If we detect a reduction in arrests, we thus cannot distinguish whether this occurs because Black or Hispanic LTs discourage certain types of policing or if, say, white LTs proactively incentivize or reward these types of policing.

5.2 Balance and First Stage

Balance on expected crime. Since our identifying variation comes from the day off group an LT is assigned to, and LTs can potentially influence which day off group they are in, it is important that either the underlying crime conditions are balanced across day off groups or, at the least, that minority LTs are no more likely to select a given lower crime day off group than non-minority LTs. Given the rotating nature of days off, this assumption appears reasonable, but we can also probe it empirically.

Using the specification in Equation 3, we investigate whether days with any predicted Black or Hispanic LT look systematically different from other days in terms of *expected* crime or arrests. In order to do so, we perform permutation tests where we randomly reshuffle district-watch outcomes (e.g. arrests, crime incidents, or 911 calls) across district-watches, within a day, and estimate the effect of having a predicted Black or Hispanic LT on this reshuffled outcome. We repeat this process 1,000 times for each of 4 different reshuffling strategies: i) reshuffling outcomes within a day, ii) within a day $\times$ CPD area (to capture only nearby districts), iii) within a day $\times$ watch, and iv) within a day $\times$ watch $\times$ CPD area. We do this for 4 different outcomes — arrests, crime, 911 calls, and emergency 911 calls — yielding 4000 iterations per outcome. Figure 3 shows the distribution of these estimated effects across each outcome. In all 4 cases distributions are centered around 0, and just around 5% of iterations are significant at the 5% level. Overall, this suggests that crime is balanced across days with and without predicted Black and Hispanic LTs.

Predicted Team composition. Another aspect that could *potentially* differ for teams managed by Black and Hispanic LTs is the composition of subordinate officers. The composition of these teams could vary for two broad reasons. First, certain types of officers could potentially be more likely to be in day off groups that have greater overlap with Black and Hispanic LTs. Officers who belong to the same day off group or adjacent day off groups work together more frequently than those in different day off groups. Second, further differences in who *actually* works on a given shift could arise for several reasons, such as transfers, selective work attendance through the use of leave or injuries, day off group re-assignments, et cetera. Here, we focus on ruling out the first scenario. That is, we show that the predicted composition of officers, based on initial day off group assignments, does not differ on average between days when a Black or Hispanic LT is scheduled to work and days when they are off.[31]

[30] Within district-watch-years for which our indicator for any Black or Hispanic LT predicted to be on duty varies, we can look at who is actually on duty on the 33% of days when we have no predicted Black or Hispanic LT. We observe at least one other LT is actually present 80% of the time, and in 80% of these cases all LTs observed on duty are white. In section 6.2, we also present estimates using a binary indicator for any predicted white LT and find roughly equal and opposite results.

[31] Since watch operations LTs are tasked with ensuring a balanced distribution of officers among teams and are thus in charge of dealing with scheduling issues that arise, any differences in the actual realized composition of officers on a particular day could be a potential mechanism for how Black and Hispanic leaders affect teams. In practice, in Section 7, we show that actual composition looks similar to predicted composition. Hence any such effects are at most small. Directly controlling for actual composition

In Table 2, we explore the correlation between having a predicted Black or Hispanic LT and predicted officer and sergeant demographics (where predictions are formed using these officers' day off groups). Imbalances here would suggest that certain officers are more likely to be sorted into the same or adjacent day off groups than others. In panel A, we see that Black and Hispanic officers and sergeants are no more likely to be in the same day off group as the Black or Hispanic LTs than other officers. Officers' total yearly arrest and use of force records from two years prior to the current year also appear statistically comparable across days when Black or Hispanic LTs are predicted to be working.

Panels B and C separate results by sergeant and officers. Focusing on officers, we do not observe significant differences when Black or Hispanic LTs are predicted to be working. However, when we examine sergeants in panel C, we identify a statistically significant relationship between having a predicted Black or Hispanic LT and the share of sergeants who are Black. We do not observe differences along other non-racial dimensions of sergeant characteristics. Out of the 24 team characteristics we study, the only two that are statistically significant pertain to sergeant race. Since one of the roles of the watch operations LT is to designate a sergeant to serve as the district station supervisor, this suggests that LTs might be handpicking some sergeants. As such, it is perhaps safest to interpret our effects as the effect of a specific leadership team (LT and sergeant), as opposed to solely an effect of the LT, and to consider the selective pairing with certain sergeants as one mechanism through which LTs exert influence. We discuss the role of sergeants in more detail in Section 6.

Altogether, while some Black or Hispanic LTs may hand-pick a sergeant to work on the same four days as themselves, there is little evidence that specific types of patrol officers are differentially pre-assigned to the same day off groups as Black or Hispanic LTs.

First stage. Finally, even though we focus primarily on the reduced form effect of having a predicted Black or Hispanic LT, it is important to understand the effective first stage. Specifically, we ask how often, between February and December, we observe a Black or Hispanic LT on duty when predicted?

Table A.7 reports the first stage coefficient from estimating Equation 3 using 'Any Black or Hispanic LT' actually present on duty as the outcome. The first stage is 0.44 (F-statistic of 259). This coefficient is virtually identical if we restrict to the teams that provide our key identifying variation (teams with at least one, but not all Black/Hispanic LTs). Among these teams, on days when we predict no Black or Hispanic LT to be present, one is nevertheless observed on duty 14% of the time. Conversely, on days when we predict a Black or Hispanic LT to be present, one is present 58% of the time. Thus, when the schedule predicts a Black or Hispanic LT to be present (i.e. does not have a scheduled day off), the probability that a Black or Hispanic LT is actually present increases by roughly 44 percentage points (from 14% to 58%). This is a strong and meaningful first stage.

As another point of comparison, consider repeatedly assigning each Black or Hispanic LT in these district-watch-years to a *randomly* selected day off group, re-constructing the indicator for any predicted Black or Hispanic LT, and re-estimating the first stage. On average, across 2000 trials, this yields a first stage of zero

further confirms that this mechanism does not appear to mediate our findings.

(or more precisely -0.0007, se = 0.0275). Intuitively, for some random day off groups this lines up with the truth and yields a positive coefficient, for others (i.e. the white LTs day off group) this yields a negative coefficient, and it yields close to null effects for the remainder. Indeed, we later leverage a randomization inference (and placebo) check in which we do precisely this.

Of course, the first stage is less than one. As discussed in Section 2.2, there are many reasons why predicted schedules will not correspond with actual work days. The two most important among them are that, i) in addition to scheduled days off, LTs also take vacations, sick leave, and can be absent for training or other assignments, and ii) LTs can move shifts or change positions mid-year.[32] These sources of error are apparent in simple cross-tabulations: among teams providing our identifying variation, we correctly identify that there is no Black/Hispanic LT on duty on 85% of days when we predict none to be (we are incorrect on 15% of these days). In contrast, and consistent with absences for other reasons and potential moves, on 58% of days when we predict a Black/Hispanic LT, this prediction is correct (and incorrect on 42%). [33]

'Compliance' with the rotation schedule does not differ by LT race within the teams that provide our identifying variation. Figure A.4 shows that our predictive accuracy does not vary by LT race. Specifically, we take all the LTs from these January teams and follow them over the course of the year (regardless of whether they move teams). We cannot reject equal predictability, or 'compliance', across any of the three statistics plotted in the figure: i) day off group predictability, conditional on staying in one's district-watch, ii) unconditional day off group predictability, and iii) actually being on duty and in the same district-watch as January.

Table A.8 further assesses whether predicted Black/Hispanic LT days are associated with differential LT substitution patterns or LT absences. We see no relationship between predicted Black/Hispanic leadership and (i) the probability that no LT is present, or (ii) the presence of an LT outside the baseline rotation.

[32] While one could re-run predictions mid-year, we prefer to use our January based predictions, since these are when day off groups first go into effect for the year, and mid-year changes could be more endogenous. For the majority of LTs who stay long-term in their shifts, this also allows us to recover longer-term effects over the course of the year, which could be larger than immediate effects.

[33] To connect these numbers to the first stage, it is useful to express the four prediction-realization cells as fractions of all days. Specifically, on 0.283 of days a Black/Hispanic LT is neither predicted nor actually on duty, on 0.049 of days a Black/Hispanic LT is not predicted to be on duty but actually is on duty, on 0.284 of days we predict a Black/Hispanic LT to be on duty but one is not actually on duty, and on 0.384 of days a Black/Hispanic LT is both predicted and actually on duty. In the simple case without additional controls, the first-stage coefficient then corresponds to the difference in conditional probabilities:

$$\beta \; = \; E(\text{Black/Hisp. LT on Duty} \mid \text{Predicted on Duty} = 1) - E(\text{Black/Hisp. LT on Duty} \mid \text{Predicted on Duty} = 0)$$

$$\approx \frac{0.384}{0.384 + 0.284} - \frac{0.049}{0.049 + 0.283} = 0.575 - 0.150 = 0.425.$$

As can be seen, since this is close to 0.44, accounting for controls changes this relatively little in practice.

6 Results

6.1 Do police teams perform differently under Black and Hispanic leaders?

Fewer Non-Index Arrests under Black and Hispanic Leaders. Table 3 presents regression estimates of Equation 3 on arrest outcomes, comparing daily within-district-watch outcomes on days with a predicted Black or Hispanic LT to days without. Notably, having a scheduled Black or Hispanic LT results in a statistically significant 0.13 decline in overall arrests, which is a 3.6% reduction relative to the mean of 3.5 daily arrests per district-watch.

Consistent with the fact that the majority of CPD arrests are of Black persons (71%), the observed reductions in arrests under Black and Hispanic LT leadership predominantly come from reductions in the arrests of Black persons. Arrests of Black persons decline by 0.10, or 81% of the total 0.13 effect on arrests. We find no statistically significant changes in the arrests of white or Hispanic persons. However, given the relative infrequency of these events, we lack the statistical precision to distinguish between true null effects and a scenario where arrests of White and Hispanic persons decline by a similar percentage of the mean as for Black persons. Specifically, the point estimates for White and Hispanic arrestees correspond to a 3.0% and 1.6% decline relative to their respective means, not far off from the 4.1% decline for Black arrestees.

In Figure 4 and Panel B of Table 3, we show that arrest reductions are concentrated in lower-severity categories. Figure 4 presents reduced-form results for arrests in each 2-digit FBI code, separately, as well as for several groupings of arrest categories. We see no statistically significant effect on index arrests, with a point estimate suggesting a 1.4% decline relative to the mean. The corresponding individual FBI codes that comprise index arrests are shown in red and also suggest muted effects. In contrast, non-index arrests decline by a highly significant 0.119, or 4% of the mean, and what we term public order offenses (colored in green in Figure 4) — liquor laws, disorderly conduct, miscellaneous non-index offenses and municipal code violations, and traffic violations — decline by 0.072 (7% of the mean, p-val < 0.001). Even as a percentage of the mean, the latter is statistically distinguishable from the 1.4% decline in index arrests,[34] implying that Black and Hispanic LTs affect *specific* types of arrests (both in levels and rates). Turning to the underlying FBI codes, we see reductions in almost all public order offenses — we estimate a 0.02 (13%, p-val = 0.011) reduction in disorderly conduct arrests, a 0.02 (5%, p-val = 0.042) reduction in miscellaneous non-index offenses, a 0.01 (10%, p-val = 0.060) reduction in municipal code violations, a 0.02 (5%, p-val = 0.053) reduction in traffic violation arrests, and no change in liquor law arrests. Outside of public order arrests, the only other large change is for non-index, drug arrests, which decline by 0.04 (6%, p-val = 0.037). Together, declines in public order arrests and non-index drug arrests account for over 88% (= 0.111/0.126) of the observed reduction in arrests. Yet, these categories account for only 48% of baseline arrests, confirming that reductions are disproportionately concentrated in lower-severity categories.

In Panel B Column 4 of Table 3, we also explore how Black and Hispanic LTs affect what Dube et al. (2025) term 'discretionary arrests' —- arrests which "may be unproductive in that the officer could have taken another course of action to resolve the situation effectively (i.e., these arrests are discretionary and

[34] The p-value on a test that the 7% and 1.4% decline are equal is 0.030.

plausibly unnecessary)". Their pre-registered discretionary arrest category, which is intended to capture likely such cases without necessarily being exhaustive, contains many of the arrests for disorderly conduct that we see decline the most. Indeed, discretionary arrests (constructed following their definition) decline by 0.007, or 11% of the mean (p-val 0.068). The officer training program studied in Dube et al. (2025) reduces these arrests by 23%, so the reduced form effect of having a predicted Black or Hispanic LT on duty on this outcome is substantial.

Benchmarking Magnitudes. How should we benchmark these reduced form effects on arrests? Prior work on the effects of Black or Hispanic officers suggests a potentially informative benchmark: how many more Black officers would need to be added to obtain the same effect of a Black or Hispanic leader? Ba et al. (2021) find that Black officers make 1.93 fewer arrests per 100 shifts (or 0.019 arrests per shift) than their white counterparts. Hence the reduced form effect of a Black or Hispanic LT is equivalent to swapping out 6.5 white officers with Black officers. In our data, the typical (median) number of patrol officers per district-watch-day is 45, of which 6 (median) to 9 (average) are Black. The effect of a Black or Hispanic LT is comparable to moving from a baseline of around 20% Black officers (the average share of Black officers) to around 35% Black officers per shift — a large 0.7 standard deviation unit increase in the share Black. Achieving the same effect of a Black or Hispanic LT thus requires nearly doubling the share of Black officers. Policymakers have often advocated for increasing the racial diversity and representation in police departments as a means to help address long-standing concerns about over-policing and disparate treatment in minority communities (US Department of Justice, 2016). Our results suggest that policies that focus on retaining and promoting minorities to leadership positions may also warrant attention.

IV Effects. The reduced form effect we estimate captures the contemporaneous effect of having a Black or Hispanic LT predicted to be on duty. In practice, LTs will occasionally be absent on predicted work days. This raises the question of whether IV estimates would be more appropriate than reduced form estimates. As discussed in Section 5, an IV approach requires an exclusion restriction that having a predicted Black or Hispanic LT only affects team outcomes on days that the Black or Hispanic LT is actually present at work. This might be a reasonable approximation to reality if daily, direct face-to-face interactions (such as during roll call at the beginning of each shift) are the primary driver of daily outcomes. However, it is also possible that repeated interactions with one's primary watch operations LT lead officers to adopt certain behaviors that persist under a different one-off, substitute LT (for example, under an out-of-rotation LT from another watch). In such cases, exclusion would not be satisfied and IV effects would be too large. For this reason, we prefer to focus on the more conservative reduced form effects throughout the paper. Nevertheless, for completeness and to help readers think about the range of possible effect sizes, we report IV results in Table A.9. Scaling the reduced form by the inverse of the 0.44 first stage in Table A.7 yields a decrease of 0.29 arrests (or an 8% reduction relative to the mean).

Quicker Emergency Dispatch Times under Black/Hispanic Leaders. Making an arrest is a time-consuming process. If officers are making fewer arrests, it is worth asking what, if anything, they do with this additional time. One possibility is that officers are available to be dispatched to calls for service more quickly. Incoming calls for service are assigned by dispatchers to available patrol units. Because units cannot be assigned while occupied with other incidents or administrative tasks, the elapsed time between call receipt and dispatch partly reflects underlying officer availability and patrol-system congestion. While our data measure time to dispatch rather than total arrival response time, prior work suggests that faster police responses can improve crime clearance and victim outcomes (Vidal and Kirchmaier, 2018; Weisburd, 2021; Deangelo et al., 2023). For example, Vidal and Kirchmaier (2018) find that a one-minute reduction in response time increases crime detection by 1.7%, while Deangelo et al. (2023) find that a one-minute reduction in response reduces the likelihood that an injury occurs by 3%, particularly among female victims. More broadly, the dispatch data provide a window into how officers spend time in the field and allocate effort across different types of activities. To examine these margins, we use data from the Chicago Office of Emergency Management & Communications' PCAD system.

The PCAD system records officer assignments and activities generated through both citizen calls for service and officer-initiated events. For each event, the system records information on the event category, timestamps associated with the incident and dispatch process, responding units, and incident locations. We first analyze dispatch response time, measured as the elapsed time between when a call is received and when a CPD unit is dispatched. In constructing this measure, we exclude observations where the recorded dispatch and received timestamps are identical, which primarily correspond to self-initiated events.

Calls are categorized by priority level. Priority 1 (Immediate Dispatch) refers to calls involving an imminent threat to life, bodily injury, or major property damage/loss, or those deemed urgent under Department guidelines. Priority 2 (Rapid Dispatch) refers to calls where timely police action has the potential to affect the outcome of the incident or is otherwise required by Department guidelines. Priority 3 (Routine Dispatch) refers to calls that do not involve an imminent threat and where a reasonable delay in response is unlikely to affect the outcome.

In Table 4, we show that teams under Black and Hispanic LTs have quicker dispatch times. Average time to dispatch decreases by approximately 1% of the mean. The effect is most pronounced for high-priority calls — those requiring immediate or rapid response and for which timely public safety intervention could influence outcomes — with dispatch times for priority 1 calls decreasing by 1.5% of the mean and dispatch times for priority 2 calls by 3% of the mean. In absolute terms, these effects correspond to reductions on the order of several seconds to roughly ten seconds, depending on the call category. While modest in magnitude, these changes are consistent with the idea that reduced emphasis on low-level enforcement under some supervisors frees officer capacity, allowing units to be assigned to incoming calls more quickly. To the extent that faster dispatch also translates into faster on-scene arrival times, these changes may have meaningful implications for public safety outcomes.

In Panel B of Table 4, we examine how officers allocate their field activity across different categories of engagements recorded in the PCAD system. Broadly, officers' time in the field can be allocated across three categories: crime enforcement (engagements related to violent and property crimes), maintenance of social order (calls related to quality-of-life issues such as disturbances, narcotics, and loitering), and community engagement (citizen assistance and community-related responses, including officer-initiated interactions).[35]

[35]See Appendix Section B.2 for details on the construction of OEMC enforcement, social-order, community-assistance, and

In column (1) we estimate a precise null effect on the number of events responded to, which suggests that, at least by this measure, overall effort or activity is not different across LT race. When we look at how officers' time is allocated across the three categories, we detect a modest decline in engagements related to social order maintenance — such as narcotics, loitering, and disturbance complaints — of approximately 0.11 calls per day, or 0.64% of the mean. At the same time, we observe an increase in community engagement activities. These engagements increase by 0.035 per day, or about 3.4% of the mean. Thus, in addition to the faster dispatch times observed in Panel A, the results suggest at least some shift in how officers allocate time and attention while in the field.

No Significant Effects on Other Outcomes In Table 5, we analyze effects on other policing and crime outcomes. In Panel A, we show that we do not detect any significant effects on stops, use of force, or complaints. Stops, which like the overall volume of officer activities recorded in the OEMC dispatch data can be interpreted as a measure of policing effort, do not change significantly. The point estimate suggests a reduction of less than 1% of the mean, with a relatively tight confidence interval (the standard error is 1% of the mean). When paired with the much larger percentage reduction in non-index and public order arrests, this is suggestive evidence that overall policing effort may not differ much across LTs of different race, but the decision to convert a given interaction into an arrest does. We also fail to detect statistically significant effects of Black and Hispanic LTs on use of force or complaints. Here, however, we are unable to rule out quantitatively meaningful changes. For example, the 95% confidence interval for use of force incidents ranges from an 8% decrease relative to the mean to a 7% increase relative to the mean.

In Table 5 Panel B, we report on court outcomes as a way to probe the potential productivity of arrests (Mas, 2006; Weisburst, 2022; Rivera, 2022) and find that the share of arrests that lead to conviction does not change. Because we need follow-up time for conviction outcomes, we restrict our sample to the years 2013–2016.[36] We are interested in eventual court conviction rates, but we cannot *directly* study court conviction rates conditional on arrests because our level of observation is a team of officers working a single watch on a single day, and it is not uncommon for these teams to go an entire watch with no arrests. Instead, we estimate the effect of a Black or Hispanic LT on the number of convictions per watch arising from arrests made by their teams. We then compare this to the change in the number of arrests per watch, including those that do not result in a conviction. Any change here could arise either because the distribution of arrests changes (e.g. if the share of index arrests increases and these are more likely to lead to conviction than non-index arrests) or because the quality of arrests changes (i.e. the conviction *rate* conditional on arrest).

We find no clear evidence that the share of arrests that end in conviction changes under Black and Hispanic LTs. This might come as a surprise given that Black and Hispanic LTs do reduce arrests for non-index arrests disproportionately and that the index arrests are generally somewhat more likely to end up in conviction on average (0.77 versus 0.53). Yet, the vast majority of overall convictions (over 80%) still come from non-index arrests. Table A.10 unpacks these two dimensions. Table A.10 panel B shows that the share of index

dispatch-time outcomes.

[36] 90% of cases in Cook County result in a sentence within 2 years and 95% are sentenced within 2.8 years. Our Court data is as of June 2019. Restricting our sample to prior to December 2016 thus gives us at least 2.5 years to observe how arrests are handled by the court.

convictions (index convictions over all arrests) indeed increases by 0.007 or 6% of the mean and the share of public order convictions (public order convictions over all arrests) decreases. This is consistent with the change in the composition of arrests under Black and Hispanic LTs leading to a higher share of index convictions and a lower share of public order convictions. When we look within category, however, in Panel C, we find little evidence that Black and Hispanic leaders affect the quality of arrests. Here the denominator is only arrests of the type considered, rather than all arrests, ruling out distributional effects. Taken together, these results suggest that while Black and Hispanic LTs supervise teams that make fewer non-index arrests, the marginal arrests they reduce within these arrest categories are of average 'quality.'

Finally, in Table 5 Panel C we show that there is no offsetting increase in crime on days when Black and Hispanic LTs are working. Given the rotational nature of the day-off calendar, large immediate crime effects would generally be unlikely. Nevertheless, if officers under Black and Hispanic LTs take systematically different visible actions and offenders respond to those actions, short-run effects on crime remain possible. We estimate a marginal decrease in all crimes (0.7%). Some of these declines, however, are likely directly attributed to the fewer arrests made under Black and Hispanic LTs, as arrests can mechanically lead to an increase in reported crimes. To address this concern, we assess the effects on crimes not associated with cleared arrests and indeed find no effect on this outcome (a 0.3% and statistically insignificant reduction). Of course, we note that our empirical strategy, with its daily variation, does not allow us to measure any potential cumulative or long-term effects of more Black and Hispanic leaders, which remains an important open question.

6.2 Alternate Specifications and Robustness

Before discussing the takeaways from these results, we present results from several alternative specifications that speak to the key role of watch operations LT specifically, our decision to pool Black and Hispanic LTs, the role of team composition and sergeants, and heterogeneity by past exposure to a minority LT and by neighborhood. We also report results from several additional robustness checks.

First, in our baseline, we pool field and watch operations LTs because field LTs who work alone will serve as watch operations LTs. An important question, however, is whether the effects are driven primarily by watch operations LTs. Several supplemental specifications suggest they are. First, in Table A.11, we investigate sole leadership by altering our indicator to be an indicator for there being only Black and Hispanic LTs predicted to be present (together or alone), effectively ensuring that we are comparing days when a Black or Hispanic LT is a watch operations LT to days when they are either not predicted present or predicted to be present together with a white LT (in which case they may or may not be acting as the watch operations LT). We find effects that are only a little smaller than the baseline, suggesting that the majority of our effect indeed arises on days when only Black or Hispanic LTs are present and hence acting as watch operations LTs. Table A.12 further confirms that we see fewer arrests both on days when we predict one out of two LTs is Black or Hispanic (relative to days with no predicted Black or Hispanic LT) *and* on days when we predict *only* Black or Hispanic LTs are present (again relative to days with no predicted Black or Hispanic LT). Effect sizes, however, are generally larger for the latter case, when we are sure that a Black or Hispanic

LT is predicted to be working as a watch operations LT. Finally, in Table A.13 we explore how effects vary when considering these two roles separately based on their predominant role in January. Effects are larger and only significant for those LTs who acted predominantly as watch operations LTs in January. Moreover, effects differ significantly from those of field LTs for overall arrests, non-index arrests, and public order arrests. Taken together, the evidence strongly points to watch operations LTs being the key driver of our effects.

Second, we probe our decision to pool Black and Hispanic LTs. We pool Black and Hispanic LTs because we cannot reject equality of their effects and pooling increases statistical power. Table A.14 disaggregates Black and Hispanic LT effects for 3 different specifications: our baseline arrests results (Panel A), and then for two specifications that try to isolate cases where the LT is working as a watch operations LT. In Panel B, we report results from the relative fixed effect specification from Section 4.3, which isolates LT effects on days they are predicted to work alone. In Panel C, we report results from the sole leadership specification discussed above for which the indicator 'Black LT' now means that there are only Black LTs predicted to be working (either alone or together), ensuring a Black LT is predicted to act as watch operations LT. Across all three specifications and all 8 arrest outcomes, we never reject equality of Black and Hispanic coefficients. In Panel A, the point estimates are larger for Hispanic LTs, but in Panels B and C, which better isolate the effect of a Black watch operations LT, effect sizes are generally larger for Black LTs. Nevertheless, these results are simply too imprecise to come to firm conclusions about differences between Black and Hispanic LTs and are ultimately consistent with the decision to pool.[37] Consistent with most arrests at baseline being of Black persons and with the qualitative findings in Ba et al. (2021), most of the estimated reductions in arrests are of Black persons regardless of whether the LT is Black or Hispanic.

Third, we confirm that effects are insensitive to controlling for the observable composition of officers. As we saw in Table 2, the predicted composition of officers across days when Black and Hispanic LTs are expected to be on duty versus days when they are not appears balanced (with the exception of Black sergeants who are more likely to share a day off group with minority LTs). Not surprisingly, then, our main findings are highly robust to the inclusion of controls for predicted officer and sergeant composition (Table A.15 Panel A). Realized team composition can nevertheless differ from predicted composition because of absences, temporary reassignments, or staffing adjustments, some of which may be influenced by watch operations LTs. However, as shown in Table A.16 Panel A, realized team composition also appears largely unrelated to predicted Black or Hispanic LT presence, much like predicted team composition. The one exception is that the share of Hispanic officers is now statistically significantly positive, though the estimate does not differ from the coefficient for predicted composition in Table 2 and is small (a 0.005 higher share off a mean of 0.19). This could be due to chance or it could be a real mechanism (if for example Black or Hispanic LTs are able to influence who fills in). Even if so, the magnitudes appear too small to meaningfully explain our results. Indeed, including (potentially 'bad') controls for actual officer and sergeant demographics and their policing histories barely moves our estimates (see Table A.15 Panel B).

[37] Effects are, on occasion, more precisely estimated for Hispanic LTs, but we estimate statistically significant reductions in public order arrests for each race across almost all specifications.

What role do sergeants play in our estimates? We have seen that predicted sergeant race is correlated with having a Black or Hispanic LT predicted on duty (Table 2). This is true even based on who actually shows up on duty (Table A.16). Teams under a predicted Black or Hispanic LT have 1.9pp higher share Black sergeants (off a base of around 10 percent). In the Dallas Police Department, Smith (2025) shows that sergeants affect team arrest outcomes. Together, these raise a question about whether our LT effects might be, in part, picking up sergeant effects. It is also independently interesting to study sergeants as they form the lower, first-line supervisory layer. In Table A.17, we study how differences in predicted sergeant composition affect team-level outcomes. First, we confirm that our baseline LT effect is unchanged by the inclusion of these predicted sergeant characteristics (including their enforcement histories), indicating that, while sergeants may still play an important role for any given LT, our effects are best interpreted as being related to the LT's characteristics and not their sergeants. Second, like Smith (2025), we see some evidence that sergeant characteristics matter for team outcomes. For example, sergeants who made more arrests as officers seem to increase team arrests. The race of the sergeant may also matter in our setting, with a 1 standard deviation (21pp) increase in the share of minority sergeants leading to 0.024 (or 0.7%) fewer arrests, but this effect is considerably smaller in magnitude than any direct LT race effect.

A final way to reinforce the idea that our results stem from the LT and not team composition is to run specifications at the officer level, incorporating officer fixed effects. Since officers work under all LTs, even if some more frequently than others, we can compare arrests for the *same officer* across days with and without a predicted Black and Hispanic LT. To do so, we construct a data set where each observation is now at the officer day-by-shift (unit by watch) level. We restrict attention to officers under the supervision of LTs who hold the titles of "Police Officer" or "Sergeant of Police" and, to avoid double-counting, we only count arrests for which the officer was the first arresting officer.[38] In Table A.18, when running regressions at the officer level (first without officer fixed effects), we find that arrests decline by 2.9% (comparable to the 3.6% decline at the district-watch level). When we disaggregate based on arrest categories, we observe reductions that are also similar to those observed at the team-level: non-index arrests decline by 3.7%, and public-order arrests decline by 5.5%. Next, we include officer fixed effects, such that we now compare arrests made by the *same* officer, across days on which a Black or Hispanic LT is predicted to be working and not. While the within-officer estimate is somewhat smaller (86–87% of the effect for overall and non-index arrests), it is statistically indistinguishable from the effect without officer fixed effects and remains highly significant. These within-officer results rule out differential officer composition as an explanation, and confirm that individual officers adjust their behavior based on which LT is supervising their shift.

Fourth, we leverage this officer level specification to ask whether officers with different past exposure to Black or Hispanic LTs are affected differently. Table A.19 compares effects at the officer level for officers with and without exposure to a Black or Hispanic LT in the prior 3 years. Prior exposure is measured using officers' attendance records. For each officer-day, we identify whether the officer worked in a district-

[38] Regular patrol officers are attached to every shift (unit and watch) they are actually observed to be working in the data. Tactical officers require some adjustment since these officers' are assigned day or night shifts that do not perfectly fit into the standard three-watch schedules of district operations and that alternate over the course of the year. To address this, we assign each tactical team officer's work time to be one of the three watches that overlap the most with the day's modal start and end times within that officer's sub-team.

watch with at least one Black or Hispanic LT on duty. We then aggregate these officer-day measures to the officer-year level and define prior exposure as an indicator for whether the officer had any such exposure over the previous three years. Of course, this single variable will capture a bundle of differences. Officers with prior exposure may differ systematically in ways that independently attenuate supervisory influence — for example, they may be more senior, more likely to have spent time in high-minority districts, or more likely to have remained in the same district-watch assignment where local enforcement norms are already well-established. Regardless, Table A.19 shows suggestive evidence that our main effects come primarily from officers without prior exposure to Black or Hispanic LTs. This suggests that even though we estimate contemporaneous effects of leadership, there may be long-lasting effects of this type of exposure. Table A.19 panel B probes robustness to adding additional interactions between predicted Black or Hispanic LT assignment and four officer characteristics that may confound the exposure measure: officer race (Black or Hispanic), above-median tenure, whether district-watch assignment over the prior three years remained the same, and above-median assignment to a high-minority district. The coefficient on the main prior-exposure interaction in this panel therefore captures the attenuation effect conditional on these alternative channels. We note also that the main effect of predicted Black or Hispanic LT assignment is now identified off a different group than the above panel — officers who are white, lower-tenure, less stable in assignment, and in lower-minority districts. We continue to find suggestive evidence of attenuation based on prior exposure even conditional on these additional differences. These results hint that in addition to the contemporaneous effects, minority LTs may have persistent impacts on officers who serve under different supervisors in the future. However, a different identification strategy would be required to causally estimate the long-run impacts of exposure to minority LTs.

Fifth, we investigate heterogeneity across neighborhoods. While Black and Hispanic LTs tend to work in higher minority-share neighborhoods, there is still variation across the types of districts they work in. In Table A.20 we explore whether the effects of a predicted Black or Hispanic LT differ across neighborhoods with a higher share Black or Hispanic, higher pre-period crime, or higher pre-period arrests. In general, we are unable to detect statistically significant differences across neighborhoods, though point estimates suggest that the arrest-reducing effects of Black and Hispanic LTs may be larger in neighborhoods with a higher minority share.

Finally, we briefly discuss several additional robustness checks: Table A.21 shows that restricting to the 128 district-watch-years that exhibit variation in having a predicted Black or Hispanic LT yields statistically indistinguishable results from our baseline. Among these same 128 district-watch-years, Table A.22 shows that replacing the indicator for having any predicted Black or Hispanic field or watch operations LT with an indicator for any predicted white field or watch operations LT yields almost equal but opposite coefficients. This confirms that the most relevant counterfactual for when a Black or Hispanic LT has a day off in this sample is having a white LT predicted on duty. Table A.23 considers alternative inference procedures, including clustering at the district-watch level (instead of district-watch-year) and a randomization inference procedure that randomly re-assigns the dummy for predicted Black or Hispanic LT within district-watch-years for which it varies.[39] Because arrests are count variables, we also show that results are similar when using a Poisson pseudo-maximum-likelihood estimator in Table A.24.

In Figure A.6, we assess whether results are driven by outlier district-watch-years (and, hence, in turn, by a few select LTs). Specifically, we estimate treatment effects of any Black or Hispanic LT specific to each district-watch-year for which this indicator varies (by including interactions between the dummy for any Black or Hispanic LT and district-watch-year dummies in Equation 3). To account for noise, we shrink the estimated treatment effects using empirical Bayes. The distribution of district-watch-year-specific treatment effects is centered around our main treatment effect estimate of -0.126 and makes clear that this average effect is not driven by outliers. We do note, however, that there is dispersion in estimated treatment effects. While almost half of district-watch-years have a treatment effect less than or equal to -0.126, one quarter of district-watch-years have a positive treatment effect larger than 0.126. This indicates that while many Black or Hispanic LTs generate reductions in arrests, this is certainly not true for all such LTs. Put another way, although certain practices, behaviors or policies appear more common among Black or Hispanic LTs, on average, there is considerable overlap across LTs of all races.

6.3 The Role of Other LT Characteristics Besides Race

Our results suggest that despite operating in the same team under the same crime reduction strategies, LTs of different races place different weights on the various possible enforcement outcomes. Black and Hispanic LTs appear to put less emphasis on proactive policing that leads to low-level arrests, reducing low-level arrests and freeing up police for faster dispatch times. These differences across race are best thought of as capturing a bundle of differences (backgrounds, lived experiences, et cetera). Here we assess the extent to which race is proxying for other *observable* characteristics and study how other LT characteristics, besides race, affect team outcomes.

In Table 6, we add other characteristics, beyond gender and race, of the LTs predicted to be on duty. Specifically, we construct binary versions of the LT characteristics studied in Section 4 (e.g. any predicted LT with more than 20 years of tenure) and include them alongside LT race and gender in our specification. Most of these other differences in observable LT characteristics are insignificantly related to outcomes. For example, we do not detect any differences in arrest patterns across gender. We also do not see differences across LTs who were promoted more quickly, despite the fact that Black and Hispanic LTs tend to have around 2 extra years of front-line officer experience prior to promotion (Table 6). We do find, however, that LTs who, as officers, had high rates of public order arrests tend to oversee teams that make more non-index and public order arrests (even conditional on race). This reinforces the idea that an LT's prior experiences and preferences (beyond those captured by race) can shape their emphases over disparate objectives.

Second, we note that these additional controls do not alter, and if anything strengthen, the baseline effects

[39] The randomization inference procedure yields standard errors close to those obtained when clustering at the district-watch-year level. The randomization inference is conducted by randomly assigning a day off group to the minority LT within each district-watch-year and reconstructing the indicator for any Black or Hispanic LT predicted to be on duty based on this randomly assigned day off group. The distributions of estimated reduced form effects resulting from this procedure are centered around zero for each outcome (see Figure A.5), and hence also serve as a placebo check.

of a predicted Black or Hispanic LT. The fact that the bundle of characteristics associated with race is not well captured by any of these other observables is consistent with differences in race being in part due to differences in lived experience that are otherwise difficult to proxy for (at least in this relatively limited set of observables).

In the next section, we probe how LTs from different racial backgrounds achieve divergent policing outcomes. It can often be hard to use existing data to shed light on specific management practices. We make progress on this problem by studying how LTs of different races grant overtime and awards.

7 Channels behind the Declines in Arrests: the Influence of Watch Operations Lieutenants

We have shown that Black and Hispanic watch operations LTs appear to weigh the tradeoffs inherent in policing differently and are able to achieve different enforcement outcomes from white LTs. These differences occur within the same district-watch and under the same overall crime strategy, which is set at a level above the LTs.[40] Given that watch operations LTs do not directly control officers in the field, how do they manage to shift outcomes? Here, we show that they leverage the discretion that comes with the position and differentially utilize tools under their control — arrest oversight, overtime grants, and award nominations — to incentivize particular types of officer behavior. These levers are not an exhaustive list of the many and often unobservable things LTs do, including for example communication during roll calls, but nevertheless provide valuable insight into how LTs affect outcomes by shaping officers' expectations of which arrests they expect to be reviewed favorably, which will earn overtime, and which will be recognized.

What levers do watch operations LTs use to influence their subordinates? Watch operations LTs have several ways to exert influence over officers. They hold roll calls at the start of each watch where they communicate plans and expectations. They manage scheduling, which includes granting or denying overtime and vacation requests, and can arrange for officers to change shifts or day off groups. They grant awards for commendable arrests, conduct complaint investigations, and take disciplinary action when performance deficiencies occur. Importantly, they also ultimately ensure compliance with arrest policies and procedures and must review and approve all arrests.

In extremely rare cases, LTs might not approve arrests brought into the station at all. In such cases the event would not show up in our data as an arrest. We issued a FOIA request for recorded events of 'release without charge,' which indeed are very rare: on average we see 0.053 releases without charge per day, relative to 3.529 non-released arrests (or less than 1.5% of cases). Because these outcomes are rare, the RWOC count outcomes reported in Table 7 Panel A are multiplied by 1,000, so coefficients can be interpreted as changes

[40]The formulation of crime prevention strategies occurs at the executive officer and district commander level (Kapustin et al., 2022). In partial empirical support of this point, we have already seen that total stops do not change significantly under Black and Hispanic LTs. Table A.25 further shows that stops across different subsets of beats are similar under Black/ Hispanic LTs and white LTs. Specifically, stops in high-Black, high-Hispanic, and high-white beats (defined on a district-specific basis based on whether the beat is at or above its district's 75th percentile population share for the given demographic) all decline by a statistically insignificant 1-2% of the mean.

in incidents per 1,000 district-watch-days. Given how rare these events are, even if all of these occurred exclusively under Black and Hispanic LTs (which is not the case), they could not directly explain our effect sizes. However, the threat of releasing without charge could be sufficient to disincentivize arrests in the field that the LT might not want to support. In equilibrium, an LT might not even need to do this. Nevertheless, in Table 7 panel A, we investigate how Black and Hispanic LTs change the share of releases without charge. We do not detect statistically significant differences in either the number or the share of releases without charge for overall arrests or non-index arrests across days when Black and Hispanic LTs are predicted to work. However, when we look at public order and discretionary arrests, some of the magnitudes and associated p-values (e.g. $p = 0.07$ for discretionary releases) suggest that this (stark) lever may indeed be used more often by Black and Hispanic LTs.

Beyond threatening non-approval of arrests, LT preferences over arrests could also filter down through the priorities they communicate and through how they utilize other forms of influence. For example, by granting coveted overtime or awards for marginal non-index arrests, an LT could effectively issue an endorsement of more active policing.[41] To assess whether LTs use these other forms of influence differently, we next investigate whether Black and Hispanic LTs differ in their overtime grants or use of awards.

Because of potential misuse of overtime, strict guidelines require prior watch operations LT approval before any overtime is worked, and LTs are required to approve only work deemed necessary for overtime. Hence, examining overtime approvals for an extension of an officer's tour of duty (which are often linked to completing an arrest) and how these differ by arrest type can provide insight into the types of arrests that different LTs value. In Table 7 panel B, we first show that there is no statistically significant difference in the number of officers who work overtime for an extension of their tour of duty across LTs of different races. However, when we then look at the type of arrests made in connection to the overtime work (i.e. arrests made four hours before the end of overtime), we note differences across LTs. Black and Hispanic LTs appear significantly more likely to grant overtime requests linked to index arrests and less likely to grant overtime requests linked to public order arrests. These differences are consistent with a differential emphasis on index relative to non-index (and particularly public order) arrests.[42] Furthermore, since overtime is generally sought after, these differences would likely be meaningful to the officers under their charge.

Last, we investigate how Black and Hispanic LTs differ in their award nominations for a given level of performance. While many awards are given by sergeants, LTs play an important role in nominating awards in recognition of outstanding performance. Although such awards are infrequently granted, exploring how the propensity to nominate officers for awards scales with different arrest types provides insights into the level of importance LTs attribute to these specific arrest categories. Our analysis uses award data from 2013 to 2020, focusing on nominations tied to incidents that occurred on a single day (as opposed to multiple days) and were specifically submitted by LTs and subsequently approved by the award board.

[41] Both current and former CPD officers have shared experiences where the preferences and attitudes of watch operations LTs have filtered down to officers' decision-making processes. When arrest reports fail to meet a watch operations LT's standard, the LT may reprimand the arresting officer and require them to revise the report.

[42] These overtime-linked arrests are not large enough to mechanically explain our main arrest results. We find similar results when we restrict the analysis to arrests occurring outside overtime periods. We also do not find meaningful differences in either the number of on-duty officers or total officer-hours on days led by Black or Hispanic lieutenants.

The majority of these awards are for honorable mentions or commendations linked to performance during arrests or crime incidents. While both sergeants and LTs can nominate any type of award, LTs tend to nominate proportionally more high-value awards. For instance, 1.6% of LT nominations are for the Superintendent's Honorable Mention and 10.23% are for department commendations, compared to 0.1% and 3.19% for sergeants, respectively. In contrast, a lower form of award, honorable mention, makes up 81% of LT nominations and 94% of sergeant nominations.

There are two key empirical considerations. First, once a nomination is made, it must be approved by both the district commander and the awards committee. Since we only observe awards that officers received, we do not capture nominations made by LTs that were denied at either of these approval stages. This means we are analyzing awards that have passed a certain threshold set by the chain of command. Second, unlike the supervisory relationship between sergeants and officers, which is formally assigned, LTs do not have a fixed link to the officers they oversee. As a result, we study LT nomination rates for all officers who worked on the same shift as the LT.

In Table 7 panel C we ask how the relationship between daily, team-level arrests and LT award nominations differs on days in which a Black or Hispanic LT is predicted to be working.[43] In column (1), for example, we see that on average each daily arrest increases the number of award nominations by a highly significant 0.004. This slope is similar on days on which Black and Hispanic LTs are predicted relative to days they are not. In columns (2) and (3), we see that each index arrest increases the number of award nominations by 0.016 and each non-index arrest increases the number of award nominations by 0.003. Again, this is similar across LTs of different races. The remaining columns start to reveal differences. For white LTs, each additional public order or discretionary arrest is associated with between 0.003 and 0.006 additional award nominations. This slope is statistically different for Black and Hispanic LTs. Indeed, instead of issuing more awards for these types of arrests, Black and Hispanic LTs are *less* likely to issue award nominations on days with high rates of these arrest types. These results indicate that while Black and Hispanic LTs do not differ from white LTs in their appreciation of index arrest performance, they tend to assign less importance to enforcement activities involving lower-level arrests.

Overall, these findings shed rare light on some of the tools police supervisors use to exert influence in settings where they have ample discretion but monitoring and formal incentive tools are lacking. Because these tools also cannot *directly* account for differences in arrests, they also highlight that LT influence operates through anticipatory channels rather than direct observation or control.

8 Conclusion

This paper shows that police LTs influence how their teams prioritize disparate and potentially competing tasks. We document meaningful dispersion in LT arrest effects and show that, among available observables, differences in LT effects are best predicted by LT race. Exploiting day off group-induced variation in whether a Black or Hispanic LT is predicted to be on duty within a given district and watch, we then

[43] Black and Hispanic LTs are no less likely to nominate awards *overall*, despite the reduction in arrests they generate.

find robust evidence that low-level arrests decline and dispatch times to emergency calls improve under the supervision of minority LTs. Hence, supervisors of different types appear successful at reallocating effort across competing objectives even in a setting with limited monitoring and ample officer discretion. We further investigate how they achieve different outcomes, finding that Black and Hispanic watch operations LTs are differentially less likely to recognize low-level arrests or issue overtime for them relative to their white peers. These findings suggest LTs influence outcomes by shaping expectations about what will be approved and rewarded.

Several avenues for future research could prove fruitful. First, we measure direct effects of LT leadership, but we also find suggestive evidence that exposure to different types of leadership might endure, affecting officers even under new leadership. Future work could attempt to estimate the long-run effects of exposure to different types of LTs on officers. Second, we use a limited set of observable LT characteristics. Better data on management practices and skills, perhaps obtained through interviews, diaries, or other sources (Bandiera et al., 2020), together with data on officers' perceptions of their supervisor's priorities and expectations could help researchers better pin down the skills and practices that affect team level outcomes and identify potential avenues for future interventions. There is likewise more work to be done on the question of whether and what reductions in arrests are welfare-enhancing. This would require tying the efforts of Black and Hispanic LTs to broader measures of community safety and sentiment. These both change much slowly than the CPD day off calendar, and hence would also require alternative sources of identifying variation in leadership.

While recent studies have focused on interventions intended to reduce use of force or discretionary arrests for all officers (e.g. Adger et al., 2022; Dube et al., 2025), our work suggests that interventions targeted at supervisors could be a valuable complement to these broader approaches. Our effect sizes are large, especially relative to what any single officer can do. Together with the fact that the LT and sergeant population is relatively small, there could well be scope for cost-effective policy interventions at the leadership level, such as recruiting, retention, or training policies. In highlighting the potential promise of intervening at this level, our work echoes a 2017 U.S. Department of Justice investigation into the CPD which emphasized the potential value of, and need for, improved supervisory practices (US Department of Justice, 2017).

References

Adger, Chandon, Matthew Ross, and C Sloan, "The effect of field training officers on police use of force," Technical Report, Working Paper 2022.

Adhvaryu, Achyuta, Anant Nyshadham, and Jorge Tamayo, "Managerial Quality and Productivity Dynamics," *Review of Economic Studies*, 2023, *90* (4), 1569–1607.

_ , Namrata Kala, and Anant Nyshadham, "Management and shocks to worker productivity," *Journal of Political Economy*, 2022, *130* (1), 1–47.

Agan, Amanda, Jennifer L Doleac, and Anna Harvey, "Misdemeanor prosecution," *The Quarterly Journal of Economics*, 2023, *138* (3), 1453–1505.

Agan, Amanda Y, Jennifer L Doleac, Anna Harvey, Anna Kyriazis, and Lauren R Schechter, "Prosecutorial reform and local crime rates," Technical Report, National Bureau of Economic Research 2025.

Anwar, Shamena and Hanming Fang, "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 2006, *96* (1), 127–151.

Ba, Bocar A, Dean Knox, Jonathan Mummolo, and Roman Rivera, "The role of officer race and gender in police-civilian interactions in Chicago," *Science*, 2021, *702* (February), 696–702.

_ , Nayoung Rim, and Roman Rivera, "The relationship between officer misconduct and conviction-less arrests," Technical Report, National Bureau of Economic Research 2024.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun, "CEO behavior and firm performance," *Journal of Political Economy*, 2020, *128* (4), 1325–1369.

_ , Iwan Barankay, and Imran Rasul, "Incentives for managers and inequality among workers: Evidence from a firm-level experiment," *The Quarterly Journal of Economics*, 2007, *122* (2), 729–773.

Bertrand, Marianne and Antoinette Schoar, "Managing with style: The effect of managers on firm policies," *The Quarterly journal of economics*, 2003, *118* (4), 1169–1208.

Best, Michael, Jonas Hjort, and David Szakonyi, "Individuals and Organizations as Sources of State," *American Economic Review*, 2023, *113* (8), 2121–2167.

Bindler, Anna and Randi Hjalmarsson, "Women in policing, gender norms and police decisions," 2023.

Bollman, Katie, Ariel Gómez, Matthew Ross, and CarlyWill Sloan, "More with Less: The Impact of Mandatory Overtime on Police Wellness and Productivity," 2025.

Bulman, George, "Law Enforcement Leaders and the Racial Composition of Arrests," *Economic Inquiry*, 2019.

Chalfin, Aaron, Benjamin Hansen, Emily K Weisburst, and Morgan C Williams Jr, "Police force size and civilian race," *American Economic Review: Insights*, 2022, *4* (2), 139–158.

Cho, Sungwoo, Felipe Gonçalves, and Emily Weisburst, "The impact of fear on police behavior and public safety," *Review of Economics and Statistics*, 2024, pp. 1–45.

Deangelo, Gregory, Marina Toger, and Sarit Weisburd, "Police Response Time and Injury Outcomes," *The Economic Journal*, 2023, *133* (May), 2147–2177.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The economics of career concerns, part II: Application to missions and accountability of government agencies," *The Review of Economic Studies*, 1999, *66* (1), 199–217.

Dixit, Avinash, "Incentives and organizations in the public sector: An interpretative review," *Journal of human resources*, 2002, pp. 696–727.

Donohue III, John, J. and Steven D. Levitt, "The Impact of Race on Police and Arrests," *Journal of Law and Economics*, 2001, *XLIV* (October), 367–394.

Dube, Oeindrila, Sandy Jo MacArthur, and Anuj K Shah, "A cognitive view of policing," *The Quarterly Journal of Economics*, 2025, *140* (1), 745–791.

Fenizia, Alessandra, "Managers and Productivity in the Public Sector," *Econometrica*, 2022, *90* (3), 1063–1084.

Ferrazares, Toshio, "Shift Structure and Cognitive Depletion: Evidence from Police Officers," Technical Report, Working Paper. UC Santa Barbara. https://ferrazares. github. io/assets ... 2025.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande, "The Personnel Economics of the State," *Handbook of Field Experiments*, 2017, (December).

Frake, Justin and Derek Harmon, "Intergenerational Transmission of Organizational Misconduct: Evidence from the Chicago Police Department," *Management Science*, 2023.

Frederiksen, Anders, Lisa B Kahn, and Fabian Lange, "Supervisors and performance management systems," *Journal of Political Economy*, 2020, *128* (6), 2123–2187.

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a self-fulfilling prophecy: Evidence from french grocery store," *Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.

Goncalves, Felipe and Steven Mello, "A few bad apples? racial bias in policing," *American Economic Review*, 2021, *111* (5), 1406–1441.

Hoekstra, Mark and Carly Will Sloan, "Does Race Matter for Police Use of Force? Evidence from 911 Calls," *American Economic Review*, 2022, *112* (3), 827–860.

Hoffman, Mitchell and Christopher T Stanton, *People, Pratices, and Productivity: a Review of New Advances in Personnel Economics*, Vol. 6, Elsevier B.V., 2025.

_ **and Steven Tadelis**, "People management skills, employee attrition, and manager rewards: An empirical analysis," *Journal of Political Economy*, 2021, *129* (1), 243–285.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 1991, *7* (special_issue), 24–52.

Janke, Katharina, Carol Propper, and Raffaella Sadun, "The Role of Top Managers in the Public Sector: Evidence from the English NHS," *NBER Working Paper*, 2020.

Jordan, Andrew, Ezra Karger, and Derek Neal, "Heterogeneous Impacts of Sentencing Decisions," 2023.

Kapustin, Max, Terrence Neumann, and Jens Ludwig, "Policing and Management," *NBER Working Paper Series*, 2022.

Kim, Taeho, "Promotion Incentives, Career Decisions, and Police Performance," 2022.

Lazear, Edward P, Kathryn L Shaw, and Christopher T Stanton, "The value of bosses," *Journal of Labor Economics*, 2015, *33* (4), 823–861.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 2006, (January).

McCrary, Justin, "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police," *The American Economic Review*, 2007, *662*, 660–662.

Mello, Steven, "More COPS, less crime," *Journal of Public Economics*, 2019, *172*, 174–200.

_ **, Matthew Ross, Stephen Ross, and Hunter Johnson**, "Diversity training and employee behavior: Evidence from the police," 2023.

Metcalfe, Robert D, Alexandre B Sollaci, and Chad Syverson, "Managers and Productivity in Retail," Technical Report, National Bureau of Economic Research 2023.

Miller, Amalia R. and Carmit Segal, "Does Temporary Affirmative Action Produce Persistent Effects? A Study of Black and Female Employment in Law Enforcement," *The Review of Economics and Statistics*, 11 2012, *94* (4), 1107–1125.

Miller, Amalia R and Carmit Segal, "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence Escalation," *Review of Economic Studies*, 2019, (February).

Minni, Virginia, "Making the invisible hand visible: Managers and the allocation of workers to jobs," *The Quarterly Journal of Economics*, 2026, p. qjag017.

Moskos, Peter, *Cop in the Hood: My Year Policing Baltimore's Eastern District*, Princeton University Press, 2008.

Owens, Emily, "The economics of policing," *Handbook of labor, human resources and population economics*, 2020, pp. 1–30.

Premkumar, Deepak, "Public Scrutiny and Police Effort : Evidence from Arrests and Crime After High-Profile Police Killings," 2021.

Rasul, Imran and Daniel Rogger, "Management of Bureaucrats And Public Service Delivery: Evidence from the Nigerian Civil Service," *The Economic Journal*, 2016, *128*, 413–446.

Rim, Nayoung, Roman Rivera, Andrea Kiss, and Bocar Ba, "The Black-White Recognition Gap in Award Nominations," *Journal of Labor Economics*, 2022.

Rivera, Roman, "The Effect of Minority Peers on Future Arrest Quantity and Quality," *SSRN Electronic Journal*, 2022.

Smith, Austin V, "The Boss in Blue: Supervisors and Police Behavior," 2025.

US Department of Justice, "The Ferguson Report: Department of Justice Investigation," *United States Department of Justice Civil Rights Division*, 2015.

_ , "Advancing Diversity in Law Enforcement Report (October 2016)," *US Department of Justice Civil Rights Division*, 2016.

_ , "Investigation of the Chicago police department," *US Department of Justice Civil Rights Division and United States Attorney's Office Northern District of Illinois*, 2017.

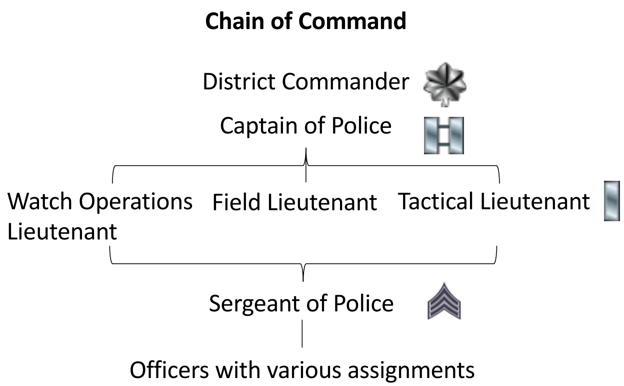
Vidal, Jordi Blanes I. and Tom Kirchmaier, "The effect of police response time on crime clearance rates," *Review of Economic Studies*, 2018, *85* (2), 855–891.

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

Weisburst, Emily K, ""Whose help is on the way?" The importance of individual police officers in law enforcement outcomes," *Journal of Human Resources*, 2022, pp. 0720–11019R2.

Figures

Figure 1: Chain of Command and Police Lieutenant Duties



Select Duties

Watch Operations LTs (staffed at all times):

1. Oversee operational and admin functions of watch
2. Ensure compliance with policy and procedures regarding arrestee processing and booking
3. Ensure watch supervisors properly assign personnel; develop and monitor strategies to manage and ensure proper oversight and response to watch personnel including, overtime, elective use of time, day-off group balance, medical roll use, beat assignment integrity
4. Designate a sergeant as district station supervisor and ensure their duties are completed
5. Respond to, assume command of, and direct supervisory action for major field incidents occurring in their district, when no field lieutenant is available.

Field LTs

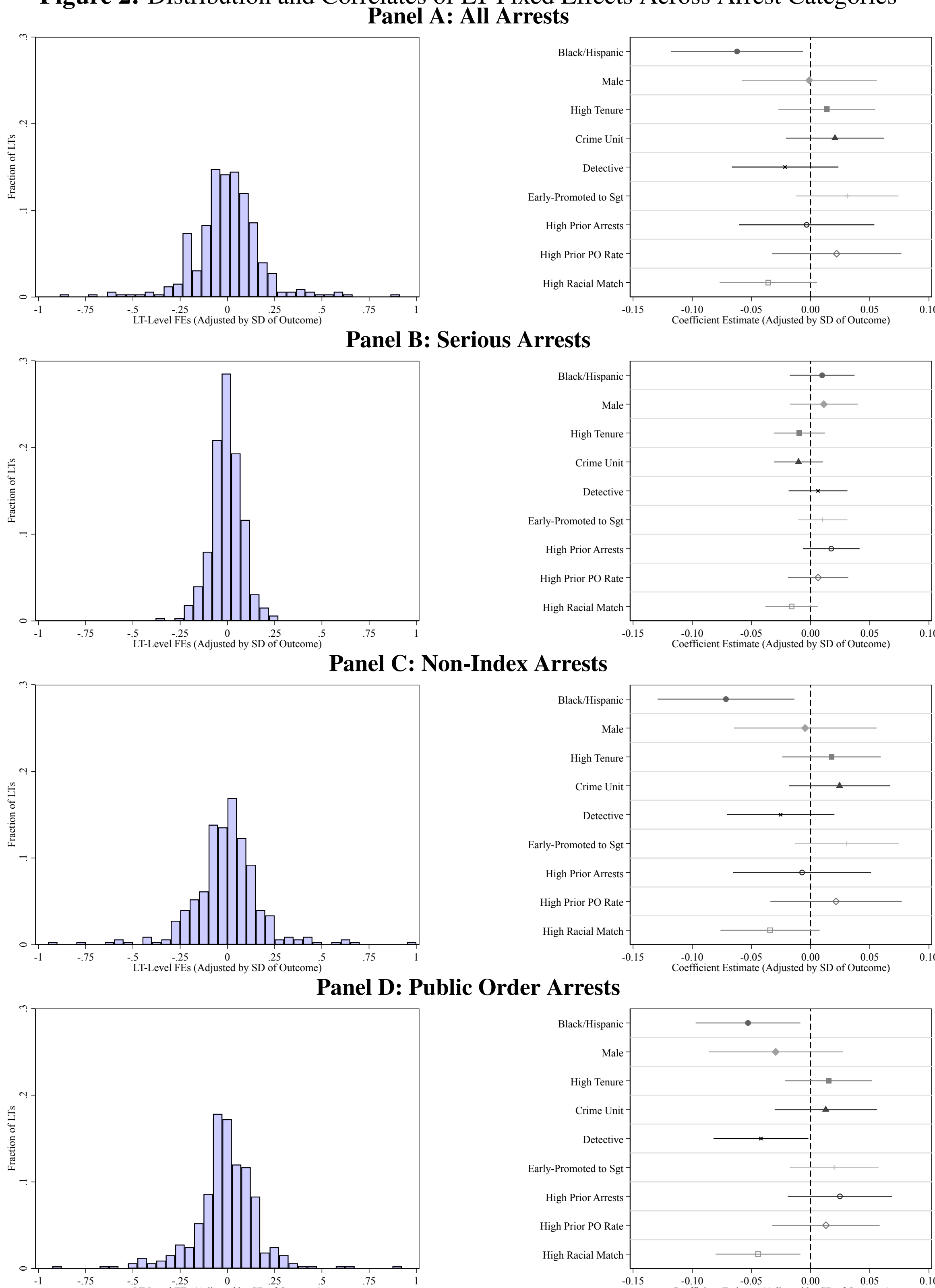
1. Respond to, assume command of, and direct supervisory action for major field incidents occurring in their district
2. Monitor behavior and performance of subordinates
3. In the absence of the watch operations lieutenant, the field or tactical lieutenant will fill the vacancy as directed by district commander or executive officer

Tactical LTs

1. Work closely with the district commander, executive officer and other supervisory personnel assigned to the district in identifying crime activity within the district
2. Be responsible for the efficiency of tactical unit operations and the coordination of tactical teams within the district
3. Ensure the watch operations lieutenant and district station supervisor are aware of daily tactical team assignments

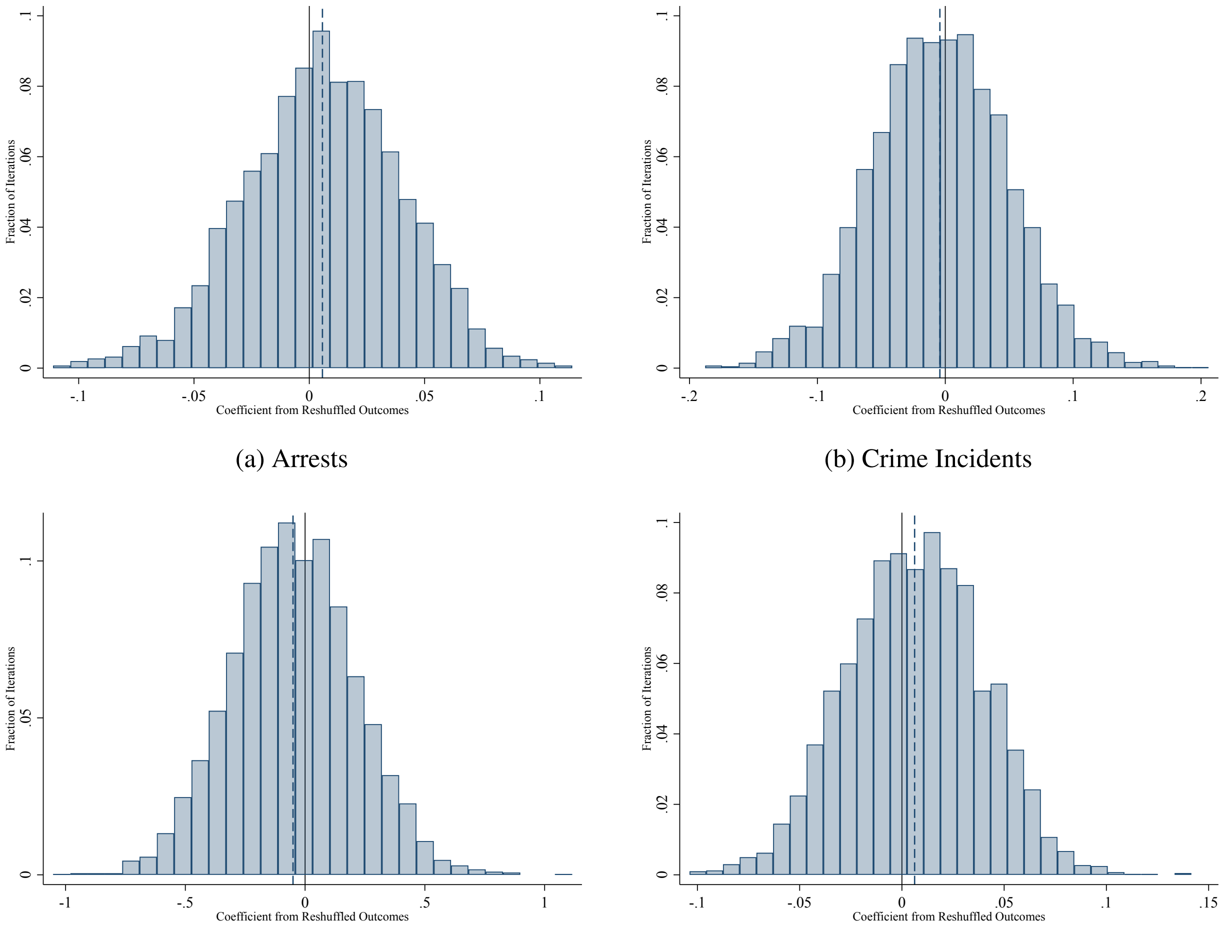
Note: This figure shows the chain of command on the left, splitting out the three types of police LTs, and the key duties of LTs on the right. Duties are taken from Special Orders S03-03-03 and S03-03-04 (https://directives.chicagopolice.org/#directive/public/6849 and https://directives.chicagopolice.org//#directive/public/6850). These directives have been revised occasionally over our study period, but roles have remained broadly consistent.

Figure 2: Distribution and Correlates of LT Fixed Effects Across Arrest Categories



Note: Left panels show the distribution of shrunken LT fixed effects from regressions of daily arrests (by category) on LT, district-watch-year, and calendar day fixed effects. The sample is restricted to LTs with at least 100 days of supervision. Right panels show coefficients from regressing LT fixed effects on LT characteristics (robust SEs; weighted by total days supervised). For variable definitions, see notes in Table A.3. High Prior PO Rate indicates a high prior public order arrest rate.

Figure 3: Permutation Tests for Balance in Predicted LT Assignment



(a) Arrests

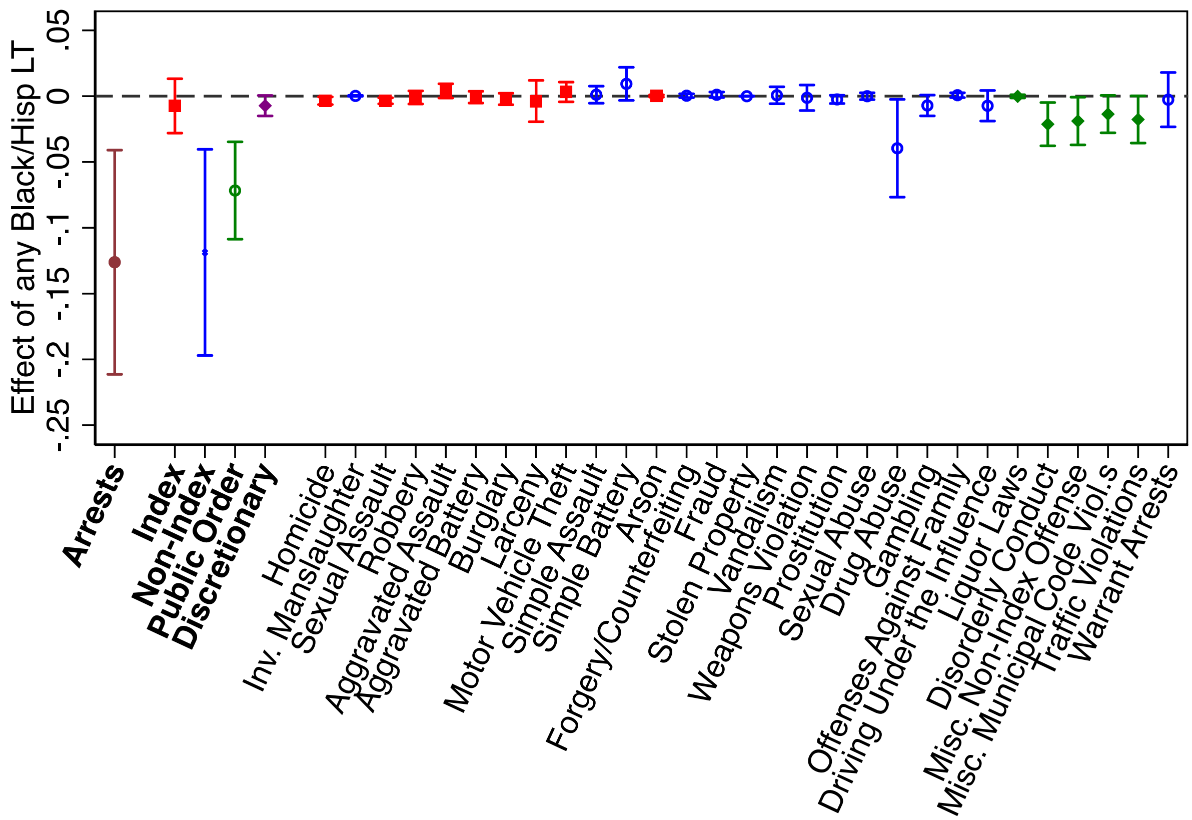
(b) Crime Incidents

(c) All 911 Calls

(d) Emergency Calls

Note: This figure reports permutation tests for balance in predicted LT assignment. In each iteration, we randomly reassign outcome realizations across district × watch teams while preserving the structure of the original data and then re-estimate the baseline specification using the reshuffled outcomes. We implement 1,000 iterations for each of four reshuffling schemes: (1) within date; (2) within date × area; (3) within date × watch; and (4) within date × area × watch. No district × watch cell is assigned to itself within an iteration. Panels display the distribution of estimated coefficients on predicted Black/Hispanic lieutenant presence across the 4,000 reshuffled samples for each outcome. The dashed vertical line indicates the mean of the permutation distribution, and the solid vertical line indicates zero. The share of permutation estimates with $p < 0.05$ is 4.6% for arrests, 4.5% for crime incidents, 5.7% for all 911 calls, and 4.7% for emergency calls; across all outcomes, the share is 4.9%, close to the nominal rate expected under random assignment.

Figure 4: Effect of a Black or Hispanic LT on Arrests in Different Categories



Note: This figure plots estimated coefficients for the reduced form effect of any predicted Black or Hispanic LT on various arrest categories. The first five entries report the estimated coefficient and 95% CIs on our 5 aggregate arrest categories: overall arrests (in brown), index arrests (red), non-index arrests (blue), public order arrests (green) and discretionary arrests (purple). Index and non-index arrests are mutually exclusive and sum to total arrests. Public order arrests are a subset of specific types of non-index arrests; discretionary arrests are also a subset of different types of arrests. The corresponding point estimates and standard errors for these first five entries can be read from Table 3 (Panel A column (1) and all of Panel B). The subsequent estimates on this figure report results for each separate 2-digit FBI code (see https://home.chicagopolice.org/statistics-data/data-requests/ for the full list of FBI codes). Effects on categories corresponding to index crimes are colored in red. Categories corresponding to non-index crimes are colored in blue or green, with public order crimes (a subset of non-index crimes) in green.

Tables

Table 1: Descriptive Statistics

	(1)
Any Pred. Black or Hisp LT	0.20
	[0.40]
Actual:	0.19
	[0.39]
Any Predicted Male LT	0.80
	[0.40]
Actual:	0.72
	[0.45]
Predicted Number of LTs	1.31
	[0.65]
Actual:	1.08
	[0.68]
Arrests	3.53
	[3.67]
Index Arrests	0.51
	[0.89]
Non-Index Arrests	3.02
	[3.41]
Public Order Arrests	1.01
	[1.69]
Discretionary Arrests	0.07
	[0.37]
N. of Team-Days	168,462
N. of District-Watch-Years	504
N. of District-Watch	63

Note: This table shows summary statistics for the daily distict-by-watch level variables used in the study. Any predicted Black/Hisp LT denotes whether a Black or Hispanic watch operations or field LT is predicted to be on duty on the given day. The number below, to the right of "Actual," denotes whether a Black or Hispanic LT actually works on the day. Male LT and number of LTs are defined analogously. We also report the total number of district-watch arrests per day, both overall and for various categories of arrests described in Section 3. Standard deviations are in brackets.

Table 2: Predicted Team Composition

	(1) Share Black	(2) Share Hisp.	(3) Share White	(4) Share Male	(5) Pred. Num.	(6) Prev. Arrests	(7) Prev. Complaints	(8) Prev. Force
	Panel A: Predicted Officer + Sergeant Characteristics							
Any Pred. Black/Hisp. LT	0.0022	0.0017	-0.0018	0.0036	0.3484	-0.0359	0.0011	-0.0020
	(0.0033)	(0.0032)	(0.0044)	(0.0031)	(0.3591)	(0.4445)	(0.0058)	(0.0087)
Y mean	0.181	0.265	0.510	0.798	55.643	41.645	0.212	0.491
	Panel B: Predicted Officer Characteristics							
Any Pred. Black/Hisp. LT	-0.0005	0.0045	-0.0018	0.0054	0.3280	-0.0400	0.0011	-0.0048
	(0.0037)	(0.0036)	(0.0048)	(0.0034)	(0.3336)	(0.4703)	(0.0062)	(0.0091)
Y mean	0.191	0.276	0.486	0.792	50.112	44.198	0.208	0.507
	Panel C: Predicted Sergeant Characteristics							
Any Pred. Black/Hisp. LT	0.0278***	-0.0198**	-0.0071	-0.0142	0.0204	-0.1605	0.0038	0.0308
	(0.0085)	(0.0098)	(0.0121)	(0.0102)	(0.0673)	(0.5783)	(0.0149)	(0.0205)
Y mean	0.087	0.166	0.720	0.849	5.531	18.500	0.246	0.345
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table examines whether district-watch-days with predicted minority LTs have different predicted team compositions, where officer-level predictions are based on their district-watch and day off group assignments in January. The specification is as in Equation 3 replacing the outcome variable with the stated variable. Panel A pertains to all officers and sergeants. Panel B pertains to just officers. Panel C pertains to just sergeants. Columns (6)-(8) examine these officers' total yearly arrests, complaints, and use of force from two years prior to the current year. *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table 3: Reduced Form: Arrests

	(1)	(2)	(3)	(4)
		Panel A: Arrests		
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.126***	-0.102***	-0.010	-0.011
	(0.043)	(0.033)	(0.008)	(0.014)
Y mean	3.529	2.501	0.326	0.670
		Panel B: Arrest Type		
	Index	Non-index	Public order	Discretionary
Any Pred. Black/Hisp. LT	-0.007	-0.119***	-0.072***	-0.007*
	(0.011)	(0.040)	(0.019)	(0.004)
Y mean	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y

Note: This table shows the reduced form results of regressing daily district-watch level outcomes on an indicator for any predicted Black or Hispanic watch operations or field LT. Panel A looks at total arrests, as well as total arrests of Black people, white people, and Hispanic people. Panel B breaks down total arrests into the four arrest categories described in Section 3. All specifications include controls for any predicted male watch or field LT, and the predicted number of watch or field LTs, as well as district-watch-year fixed effects and calendar-date fixed effects. Standard errors, clustered at the district-watch-year, are in parentheses. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table 4: Effects on Dispatch Time & Cases

	(1)	(2)	(3)	(4)
		Panel A: Time to Dispatch		
	All	Immediate	Rapid	Routine
Any Pred. Black/Hisp. LT	-.0976*	-.0778*	-.176***	-.142
	(.058)	(.0419)	(.0677)	(.0893)
Y mean	9.4	5.4	6.6	14.3
Observations	147,420	147,420	147,420	147,420
		Panel B: Types of Engagement Events		
	All	Enforcement	Social Order	Community Engagement
Any Pred. Black/Hisp. LT	-.0525	-.0543	-.113*	.0365***
	(.140)	(.0546)	(.0587)	(.014)
Y mean	65.0	16.6	17.2	1.0
Observations	168,462	168,462	168,462	168,462

Note: The data come from Chicago Emergency Management & Communications' Police Computer Aided Dispatch (PCAD) system. In Panel A, we analyze call dispatch time (in minutes), measured as the difference between the timestamp when a 911 call is received and when a CPD unit is dispatched. Calls are categorized by priority level. Priority 1 (Immediate Dispatch) refers to calls involving an imminent threat to life, bodily injury, or major property damage/loss, or those deemed urgent under Department guidelines. Priority 2 (Rapid Dispatch) refers to calls where timely police action has the potential to affect the outcome of the incident or is otherwise required by Department guidelines. Priority 3 (Routine Dispatch) refers to calls that do not involve an imminent threat and where a reasonable delay in response is unlikely to affect the outcome. If a district-watch does not receive a call of a given priority level on a particular day, we impute the dispatch time using the district-watch's daily average dispatch time across all calls. This data is recorded beginning 2014, so we restrict our analysis to 2014-2020. In Panel B, we examine the types of events officers respond to. Crime enforcement includes engagements related to violent and property crimes. Social order includes engagements related to quality-of-life issues such as disturbances, narcotics, and loitering. Community engagement includes events with descriptions related to citizen assistance and community interactions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Other Outcomes

	(1)	(2)	(3)	(4)
	PANEL A: POLICE ACTIVITY AND MISCONDUCT			
	Stops	Use of Force	Complaints	Force Complaints
Any Pred. Black/Hisp. LT	-0.1120	-0.0006	0.0009	-0.0005
	(0.1175)	(0.0037)	(0.0013)	(0.0007)
Observations	168,462	168,462	168,462	168,462
Y mean	11.4984	0.0985	0.0130	0.0038
Baseline FE and controls	Y	Y	Y	Y
	PANEL B: CHARGING OUTCOMES			
	Guilty/Arrest			
Difference in Ratio	0.007			
	[0.303]			
Observations	84,231			
Mean Ratio	0.567			
Baseline FE and controls	Y			
	PANEL C: CRIME			
	Crime	Crime w/o Arrests		
Any Pred. Black/Hisp. LT	-0.0771*	-0.0238		
	(0.0431)	(0.0390)		
Observations	168,462	168,462		
Y mean	11.0980	8.5925		
Baseline FE and controls	Y	Y		

Notes: In Panel A, we run the same regression as in Table 3 for 4 additional outcomes: total stops, total use of force incidents, total complaints, and total complaints related to use of force. For Panel B, we leverage court records to study 'guilty' ratios. Specifically, we define an arrest as leading to a conviction if the case is disposed of as either a misdemeanor or a felony. Otherwise, an arrest is defined as non-guilty (if the case is dismissed, not matched to the court data, or does not have clear case outcomes). We estimate coefficients in regressions where the dependent variables are guilty arrests and total arrests. We add the estimated coefficients for guilty and overall arrests to their respective means and compare the resulting ratio with the ratio of the means. The p-values are in brackets. See Table A.10 for more details. In Panel C, we run the same regression as in Table 3 on crime outcomes: first for all cleared and uncleared crimes, and second after removing crimes that led to arrests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effects of Black/Hispanic Lieutenants Controlling for Other LT Characteristics

	Arrests				911 Calls (2014+)						
	(1) All	(2) Serious	(3) Non-Index	(4) Public Ord.	(5) Enforce.	(6) Social Ord.	(7) Community	(8) Time Disp.	(9) Time p1	(10) Time p2	(11) Time p3
Any Pred. Black/Hispanic LT	-0.152***	-0.013	-0.138***	-0.083***	-0.055	-0.125*	0.034**	-0.097*	-0.063	-0.178***	-0.165*
	(0.047)	(0.011)	(0.043)	(0.020)	(0.056)	(0.068)	(0.015)	(0.059)	(0.042)	(0.063)	(0.092)
Any Pred. Male LT	-0.044	-0.002	-0.041	-0.013	0.059	0.105	0.005	0.093	0.076*	0.105*	0.055
	(0.039)	(0.010)	(0.037)	(0.020)	(0.063)	(0.072)	(0.012)	(0.058)	(0.043)	(0.061)	(0.090)
Any Pred. High-Tenure LT	-0.023	-0.004	-0.019	0.045**	0.007	-0.037	-0.021	0.019	-0.026	0.068	0.092
	(0.037)	(0.010)	(0.037)	(0.020)	(0.063)	(0.072)	(0.014)	(0.052)	(0.039)	(0.061)	(0.085)
Any Pred. Crime-Unit LT	0.024	0.006	0.018	0.001	-0.034	0.005	0.009	0.053	0.021	0.021	0.112
	(0.034)	(0.007)	(0.034)	(0.016)	(0.058)	(0.053)	(0.013)	(0.046)	(0.037)	(0.051)	(0.070)
Any Pred. Detective LT	0.019	0.009	0.011	-0.006	-0.023	-0.058	-0.010	-0.021	0.005	0.034	-0.039
	(0.038)	(0.009)	(0.035)	(0.017)	(0.050)	(0.057)	(0.011)	(0.050)	(0.039)	(0.058)	(0.076)
Any Pred. Early-Promoted to Sgt LT	-0.014	-0.012	-0.002	0.008	-0.018	0.078	-0.017	0.068	0.044	0.071	0.078
	(0.034)	(0.008)	(0.032)	(0.016)	(0.058)	(0.059)	(0.013)	(0.049)	(0.040)	(0.052)	(0.073)
Any Pred. High Prior Arrests LT	-0.031	-0.017*	-0.014	0.008	-0.030	-0.019	-0.003	0.012	-0.007	-0.005	0.025
	(0.040)	(0.010)	(0.037)	(0.022)	(0.057)	(0.063)	(0.014)	(0.063)	(0.045)	(0.072)	(0.091)
Any Pred. High Prior Public-Order Arrest Rate LT	0.077**	0.010	0.067*	0.038*	-0.068	-0.089	-0.008	0.022	0.008	-0.064	0.104
	(0.038)	(0.011)	(0.036)	(0.021)	(0.070)	(0.066)	(0.017)	(0.065)	(0.045)	(0.072)	(0.099)
Any Pred. Racial Match with District Population	0.047	0.008	0.038	0.016	-0.044	0.118*	0.021	-0.027	-0.051	-0.041	-0.010
	(0.042)	(0.010)	(0.041)	(0.021)	(0.055)	(0.066)	(0.016)	(0.062)	(0.049)	(0.065)	(0.091)
Y mean	3.529	0.514	3.016	1.015	15.780	16.533	1.039	9.445	5.368	6.625	14.317
Observations	168,462	168,462	168,462	168,462	147,420	147,420	147,420	147,420	147,420	147,420	147,420
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table reports the effect of predicted Black/Hispanic LTs on policing outcomes, controlling for other LT characteristics. All LT characteristics are measured as indicators for whether the team is predicted to have any LT with the given characteristic on that day. See Appendix Section B.1 for details on variable construction. Missing values for early promotion, prior arrest rates, and prior activity are set to zero, with separate missing-indicator controls included. All specifications include district-watch-year and date FE. Standard errors are clustered at the district-watch-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: How Do Lieutenants Exert Their Influence?

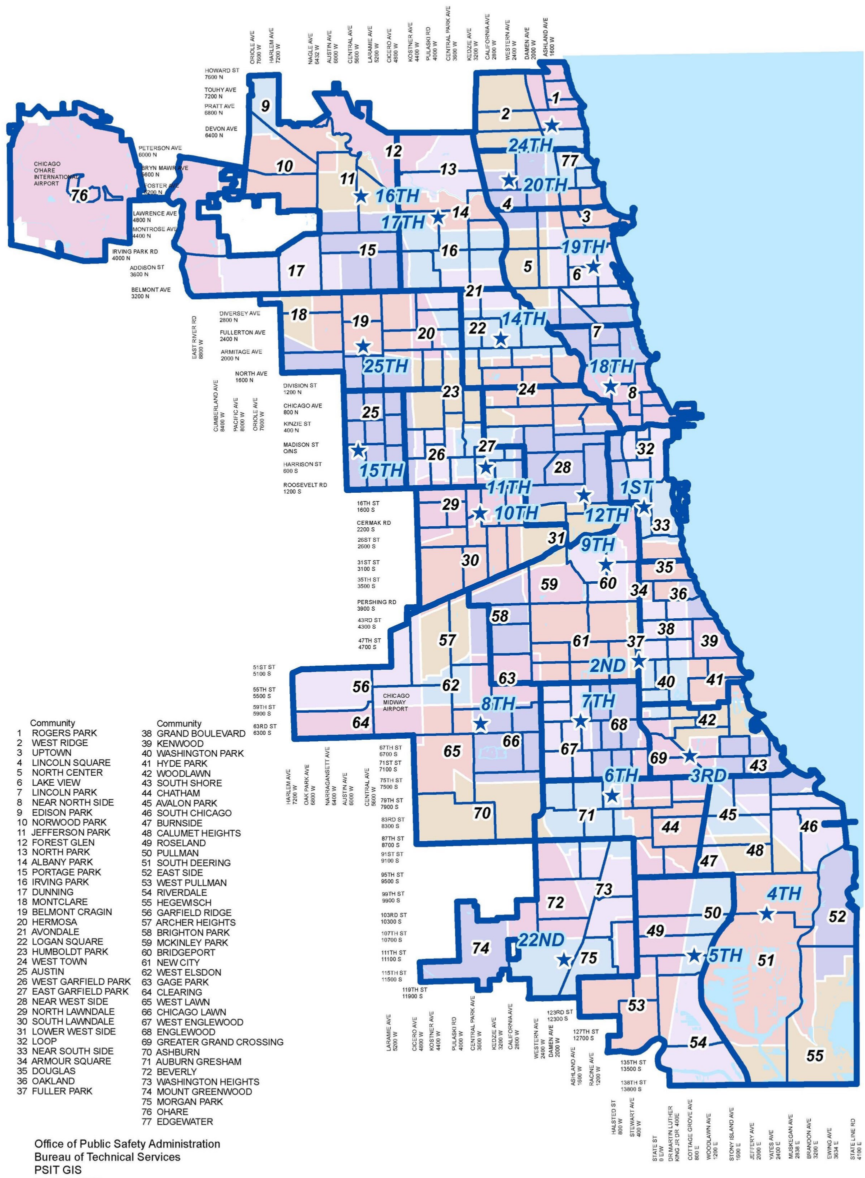
	(1)	(2)	(3)	(4)	(5)
	PANEL A: RELEASE WITHOUT CHARGE (PER 1,000 DISTRICT-WATCH-DAYS)				
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	0.073	1.668	-1.595	2.186	0.289
	(4.031)	(2.496)	(3.069)	(2.118)	(0.178)
Y mean	53.389	28.279	25.110	6.488	0.297
Difference in Ratios	0.001	0.004	-0.0002	0.003	0.005*
	[0.621]	[0.406]	[0.841]	[0.208]	[0.072]
Mean Ratio	0.015	0.055	0.008	0.006	0.004
Observations	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y
	PANEL B: ARRESTS MADE BY OVERTIME OFFICERS				
	All	Serious	Non-Index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	0.002	0.006***	-0.004	-0.005**	-0.000
	(0.007)	(0.002)	(0.006)	(0.002)	(0.001)
Y mean	.104	.029	.0745	.0237	.00236
Observations	147,420	147,420	147,420	147,420	147,420
	PANEL C: CONTRIBUTION OF ARRESTS ON LT AWARDS ARREST CATEGORIES				
	All	Index	Non-Index	Public Order	Discretionary
Baseline effect	0.004***	0.016***	0.003***	0.003*	0.006
	(0.001)	(0.003)	(0.001)	(0.001)	(0.007)
Differential effect (Black/Hisp. LT)	-0.001	-0.002	-0.001	-0.006*	-0.028**
	(0.001)	(0.006)	(0.001)	(0.003)	(0.013)
Observations	168,462	168,462	168,462	168,462	168,462
Y mean	0.049	0.049	0.049	0.049	0.049
Baseline FEs & Controls	Y	Y	Y	Y	Y

Note: This table examines three distinct tools that LTs have discretion over and that could be used to influence their subordinates: release without charge (RWOC), overtime allocation, and award nominations. Results compare outcomes on days with predicted Black or Hispanic LTs to other days. Panel A analyzes both the number of RWOC decisions and the share of arrests released without charge, overall and by arrest category. RWOC count outcomes are multiplied by 1,000, so coefficients and means can be interpreted as changes in incidents per 1,000 district-watch-days. For RWOC shares, we estimate separate regressions for RWOC decisions and total arrests, add the estimated coefficients to their respective sample means, and compare the implied RWOC-to-arrest ratio to the baseline ratio; bracketed p-values report tests of equality between these ratios. Panel B examines differences in overtime activity under predicted Black or Hispanic LTs. It analyzes arrests made by officers working overtime, overall and by arrest category. Overtime arrests are defined as arrests occurring during the overtime window or within four hours prior to its start. The smaller sample size reflects the availability of overtime data beginning in 2014. Panel C studies differences in awards given conditional on performance. Using award data from 2013–2020, we focus on single-day incidents nominated by LTs. We regress total counts of awards on arrests (by category), an indicator for a predicted Black or Hispanic LT, and their interaction. The arrest coefficient captures the relationship between arrests and awards under non-Black/Hispanic LTs, while the interaction term captures how this relationship differs under predicted Black or Hispanic LTs. All specifications include the baseline fixed effects and controls used in Table 3. Standard errors are clusted at the district-watch-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Online Appendix

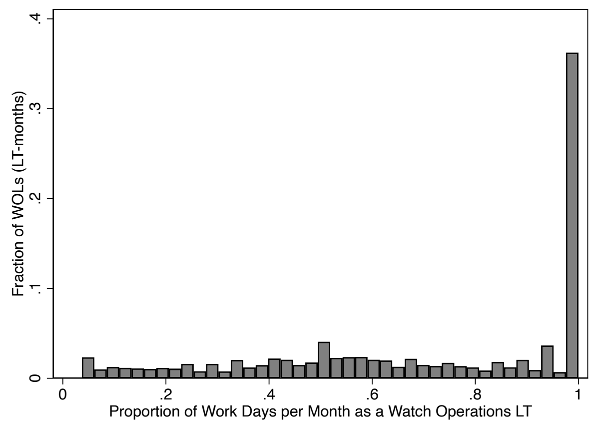
Appendix Figures

Figure A.1: Map of CPD Districts



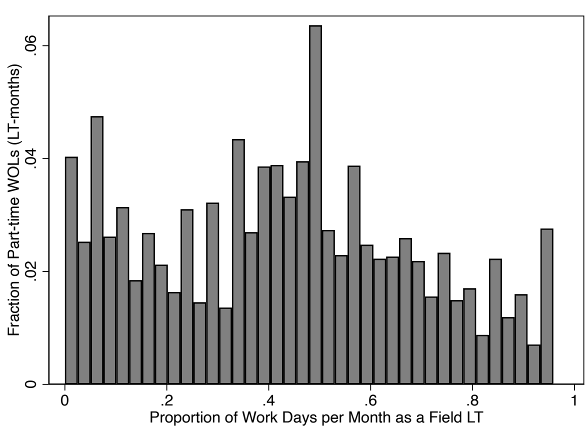
Note: This map, taken directly from https://home.chicagopolice.org/wp-content/uploads/2020/07/CPD-Area-Map.pdf, shows the 22 Chicago police districts (denoted by the solid, thicker blue lines). The district headquarters is denoted by a star. The thinner blue lines denote beats. The multi-colored areas numbered 1-77 are Chicago community areas.

Figure A.2: Interchangeability of LT Roles For Watch and Field LTs

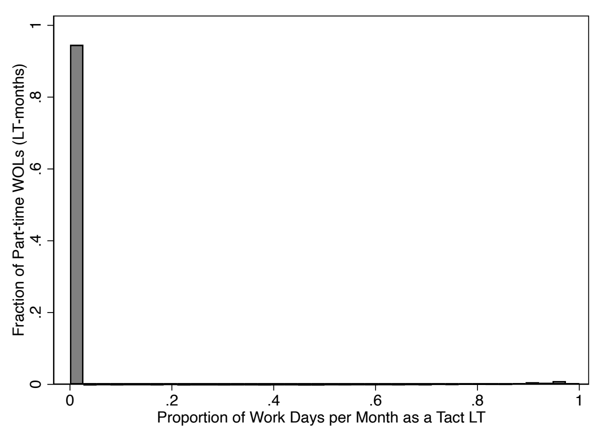


(a) Frac. of Days as WOL

Among Part-time WOLs:



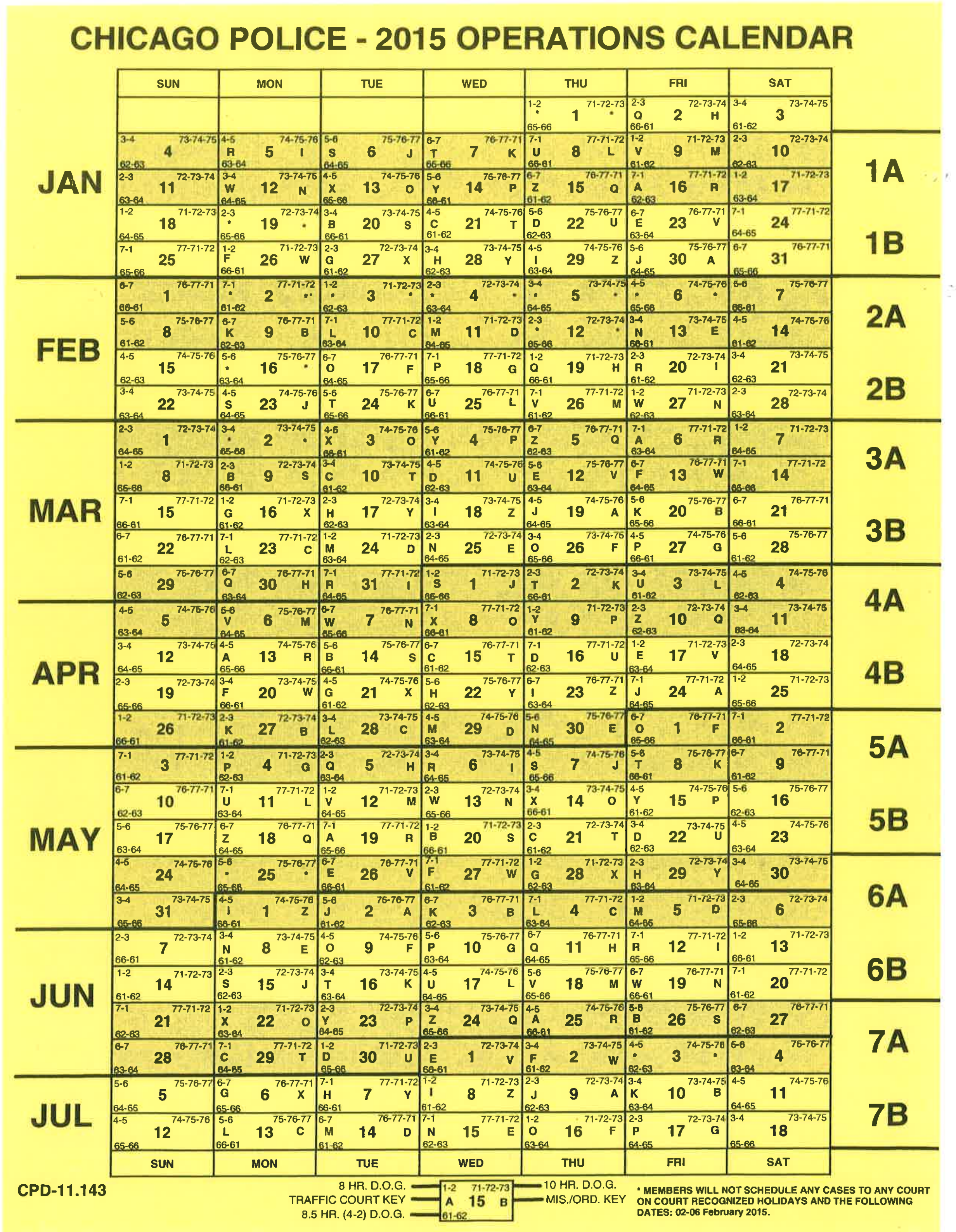
(b) Frac. of Days as field LT



(c) Frac. of Days as Tact LT

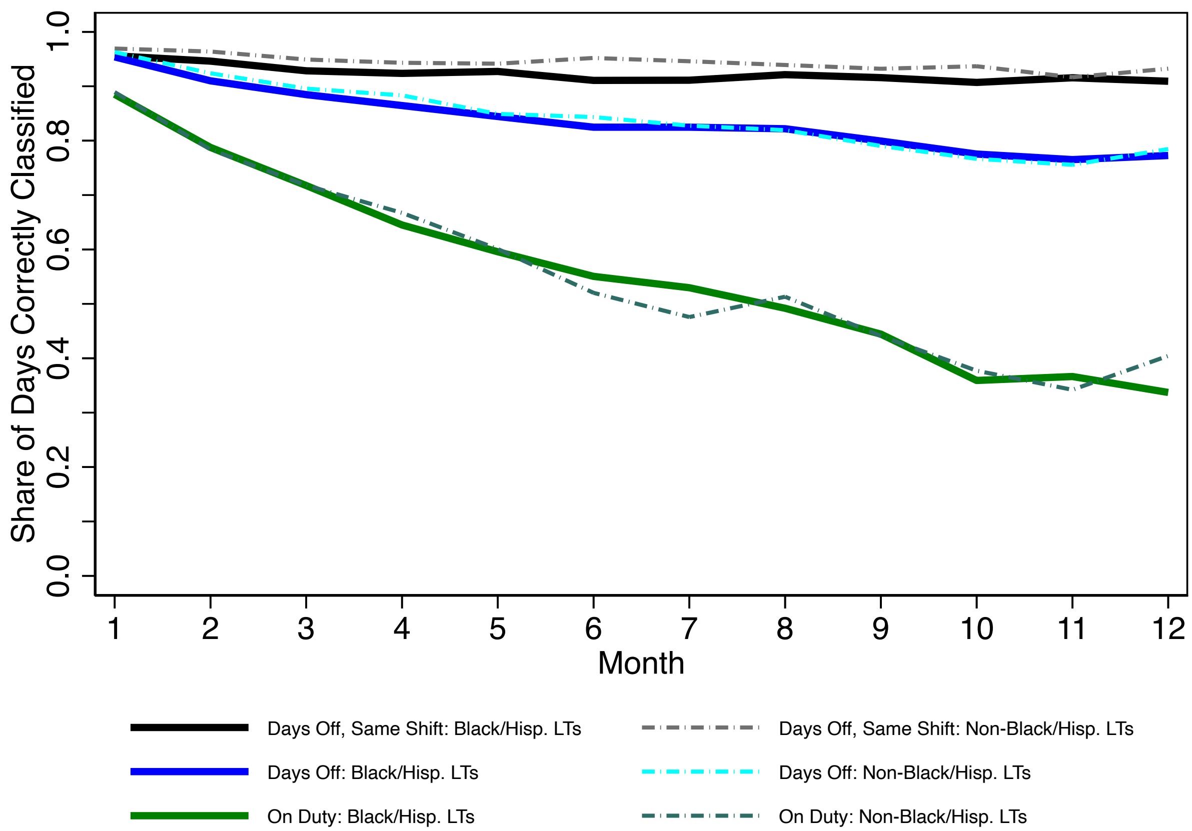
Note: This figure shows how LTs who work as watch operations LTs split their work days. In the first figure, we show distributions of work days as watch operations LTs (WOL) among LTs who ever work as WOLs in a month. Specifically, the data are at the LT-month level, and we plot, among those who serve as a WOL at least once in a month, the proportion of days spent as a WOL in that month. The second and third panels focus on LTs that serve as part-time WOL in any given month (those that worked as a WOL at least once in a month but not all the time). The second panel shows a distribution of days spent as a field LT. The third panel shows the distribution of days spent as a tactical LT.

Figure A.3: day off Group Calendar in 2015



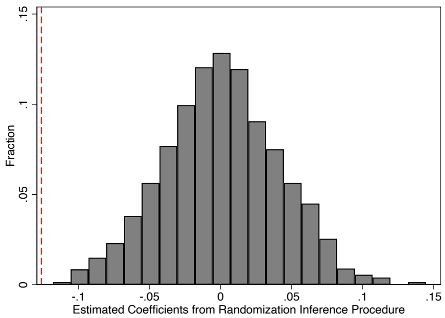
Note: This figure shows the 2015 day off group calendar obtained from the CPD. The numbers between 61 and 66 in the bottom left corner of each box indicate whether that group has a day off on the given day. For example, on Sunday Jan 4, the 62-63 indicates that day off groups 62 and 63 are not expected on duty on this day. The other numbers in the top left and right corners pertain to day off groups for specialized personnel not under consideration in this study.

Figure A.4: Predicted vs. Actual Schedule Compliance, By LT Race

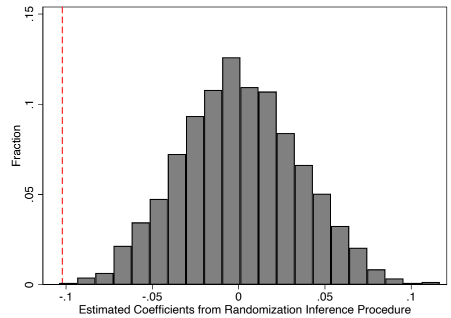


Note: For this figure, we predict days off for LTs using their January day off group assignments and district-watch assignment. Then, we merge to LTs actual recorded days off and days present on duty based on attendance records. We calculate three statistics: i) the share of days we correctly identify as a 'day off', conditional on LT-days on which LTs remain in their same district-watch, ii) the unconditional share of days we correctly identify as a 'day off', and iii) the share of days we correctly identify as having an LT on duty. We plot of each these series in each month separately for Black and Hispanic LTs and non-Black and Hispanic LTs. We restrict to LTs in January teams that provide our identifying variation (i.e. teams for which there is at least one Black or Hispanic LT in January and not all LTs are Black or Hispanic). The comparisons reveal that predictability, or 'compliance', does not vary by LT race across any of these three statistics. Simple bivariate regressions of the statistic on an indicator for a Black or Hispanic LT, pooling all months from February-December and clustering at the district-watch-year level, confirm no significant relationship between any of the three statistics and the race of the LT.

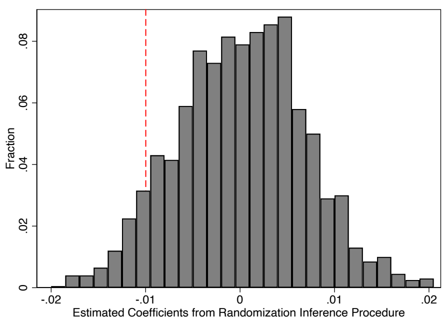
Figure A.5: Coefficients From Randomization Inference Procedure



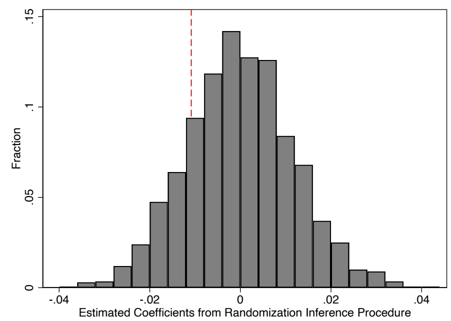
(a) Arrests



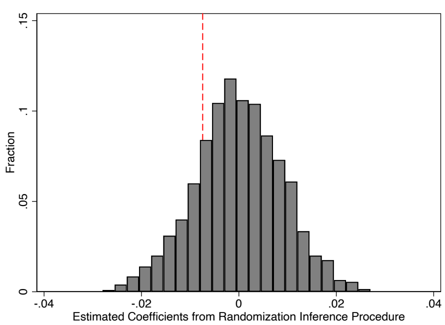
(b) Arrests of Black Persons



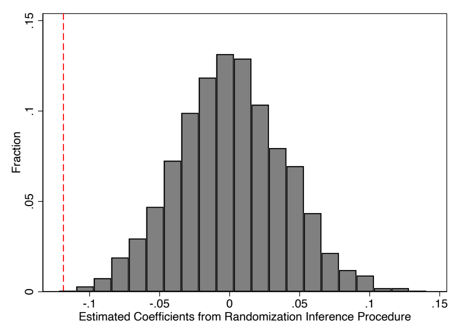
(c) Arrests of White Persons



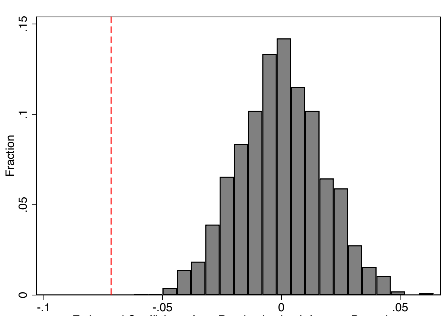
(d) Arrests of Hispanic Persons



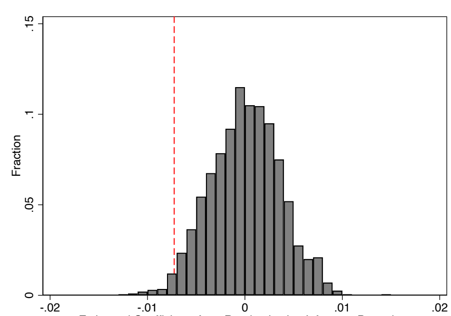
(e) Index Arrests



(f) Non-Index Arrests



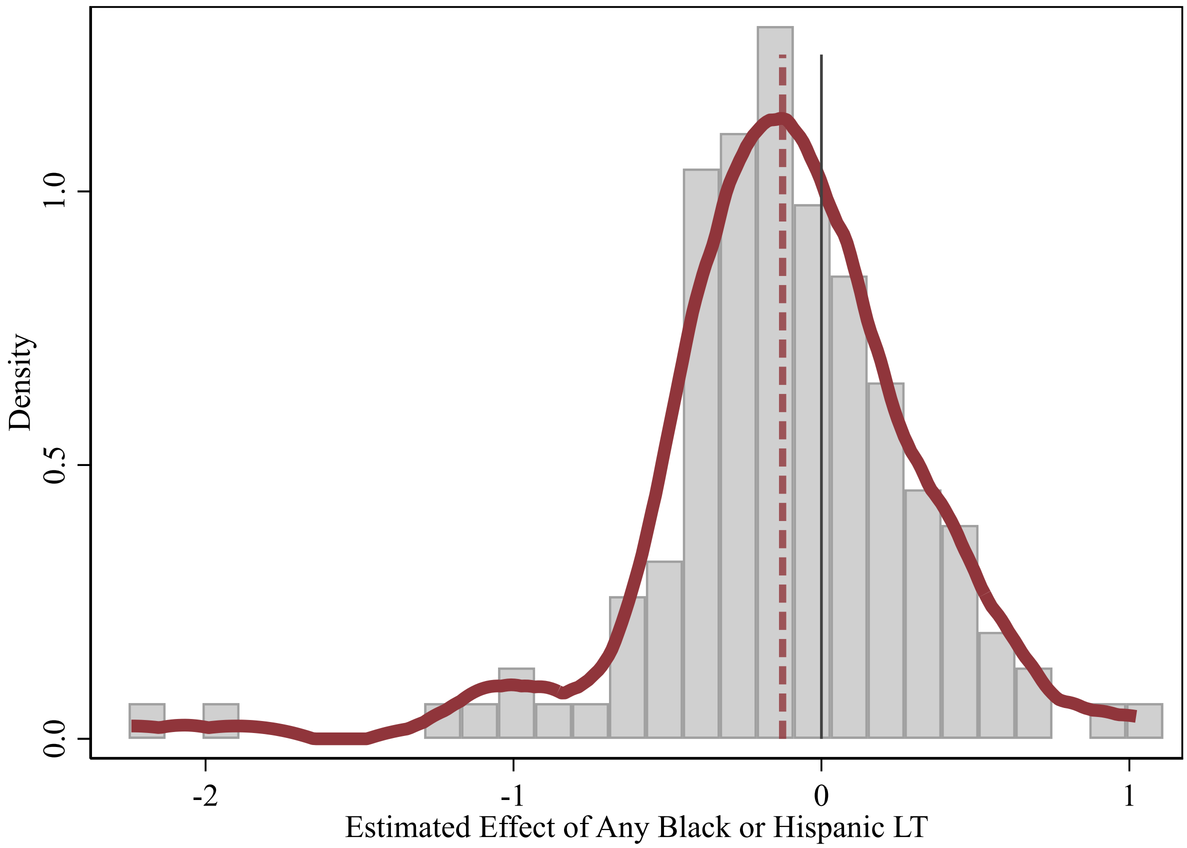
(g) Public Order Arrests



(h) Discretionary Arrests

Note: This figure plots the distribution of reduced-form coefficients that arise from performing 2,000 trials of the randomization inference procedure described in Section 6.2, in which we reassign the indicator for any Black or Hispanic LT predicted to be on duty based on randomly selected day off groups. Each subfigure corresponds to a separate outcome variable. The *actual* estimate is denoted by the red dashed line.

Figure A.6: Estimating Treatment Effects Separately for each District-watch-Year



Note: This figure plots the distribution of treatment effects from a specification in which we estimate the effects of having a Black or Hispanic LT separately for each district-watch-year (implemented by including terms for the interaction of district-watch-year and any Black or Hispanic LT predicted to be on duty). We are able to estimate 128 separate treatment effects (for the 128 district-watch-years for which any predicted Black or Hispanic LT varies; i.e. for which at least some, but not all, day off groups have a Black or Hispanic LT). Coefficients are shrunk using empirical Bayes, but the figure looks very similar without shrinkage. The histogram shows the distribution of shrunk treatment effects and the solid maroon line plots the corresponding kernel density. The distribution of estimated treatment effects is centered around -0.124 (median -0.11) arrests per day, very close to the estimated treatment effect of -0.126 from Table 3, which is indicated by the dashed maroon line.

Appendix Tables

Table A.1: Leader Characteristics

	(1) All	(2) Black/Hispanic	(3) White	(4) Difference
Male	0.81	0.75	0.83	0.09
	(0.39)	(0.44)	(0.38)	(0.04)
Appointment Year	1992.88	1992.66	1992.96	0.30
	(5.36)	(4.93)	(5.52)	(0.54)
Crime Unit	0.34	0.41	0.31	-0.10*
	(0.47)	(0.49)	(0.46)	(0.05)
Detective	0.30	0.32	0.29	-0.03
	(0.46)	(0.47)	(0.45)	(0.05)
Time to Sgt	10.52	11.89	9.96	-1.94***
	(3.66)	(4.00)	(3.36)	(0.43)
Prior Arrests	265.89	288.64	257.21	-31.43
	(378.98)	(410.84)	(366.41)	(42.46)
Prior Public-Order Arrest Rate	0.26	0.25	0.27	0.02
	(0.13)	(0.15)	(0.12)	(0.02)
N. of Lieutenants	442	122	320	442

Note: This table shows the characteristics of leaders who are active as watch operations or field LTs in our study period. See Appendix Section B.1 for details on variable construction. Standard deviations are in brackets. Column 4 reports a standard t-test for differences in means *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A.2: Noise-Corrected Signal Variance of Lieutenant Fixed Effects

	Signal SD $\sqrt{\hat{\sigma}_{\alpha}^{2}}$	SD(Outcome)	Normalized Signal SD $\sqrt{\hat{\sigma}_{\alpha}^{2}}$/SD(Outcome)	Mean
All Arrests	0.72	3.67	0.20	3.53
Index Arrests	0.08	0.89	0.09	0.51
Non-Index	0.69	3.41	0.20	3.02
Public Order	0.30	1.69	0.18	1.01

Notes: Each row reports noise-corrected signal variance estimates for LT fixed effects estimated from equation 1, following the empirical Bayes procedure of Weisburst (2022). Signal SD is $\sqrt{\hat{\sigma}_{\alpha}^{2}}$, where $\hat{\sigma}_{\alpha}^{2} = \hat{\sigma}_{v}^{2} - \hat{\sigma}_{\varepsilon}^{2}$ is the estimated variance of true LT effects after removing the contribution of sampling noise. Normalized Signal SD scales this by the standard deviation of the outcome. Sample restricted to LTs with at least 100 days of supervision.

Table A.3: Correlation between LT FEs and Characteristics

	Arrests FEs	Index FEs	Non-Index FEs	Public Order FEs	Discretionary FEs
Black/Hispanic	-.228**	.0087	-.244**	-.0892**	-.0065**
	(.104)	(.0124)	(.1)	(.038)	(.0028)
Male	-.0042	.0099	-.0157	-.0497	-3.6e-04
	(.107)	(.0129)	(.105)	(.0486)	(.0027)
High Tenure	.0504	-.0085	.0603	.0258	3.5e-04
	(.0764)	(.0097)	(.072)	(.0315)	(.0023)
Crime Unit	.0758	-.009	.0836	.0216	-7.0e-04
	(.0775)	(.0093)	(.0742)	(.0371)	(.0023)
Detective	-.0792	.0056	-.0862	-.0708**	.0011
	(.0842)	(.0112)	(.0788)	(.0344)	(.0025)
Early-Promoted to Sgt	.114	.009	.104	.0336	.0024
	(.0808)	(.0094)	(.0762)	(.0322)	(.0022)
High Prior Arrests	-.0123	.0155	-.0244	.0419	.0086***
	(.107)	(.0108)	(.101)	(.0379)	(.0025)
High Prior Public Order Arrest Rate	.0809	.0057	.0735	.0219	.0017
	(.102)	(.0115)	(.0961)	(.0389)	(.0026)
High Racial Match with District Population	-.131*	-.0142	-.117	-.0749**	-.0026
	(.0769)	(.0099)	(.0727)	(.0308)	(.0024)
Observations	430	430	430	430	430

Note: This table regresses shrunken fixed effects (FEs) for LTs who supervised teams in the sample on their individual characteristics. On each day of actual supervision, we identify the LTs assigned to watches 1-3 and, when multiple LTs are present, prioritize the Watch Operations LT. After applying Bayesian shrinkage, we weight the regression by the number of days each LT supervised. Each characteristic is entered in a separate regression. Robust standard errors are reported. The sample includes 430 lieutenants for covariates with complete coverage. Sample sizes are smaller for regressions using variables with incomplete coverage: unit histories (429), promotion timing (380), prior arrest histories (280), and prior public-order arrest histories (263). All characteristics are binary indicators equal to one if the LT exceeds the sample median, except where the variable is inherently binary. See Appendix Section B.1 for details on variable construction.

Table A.4: Correlation between LT FEs and Characteristics: All together

	(1) Arrests FEs	(2) Index FEs	(3) Non-index FEs	(4) Public Order FEs	(5) Discretionary FEs
Black/Hispanic	-.224*	.0151	-.248**	-.0907**	-.00602**
	(.122)	(.0133)	(.117)	(.0396)	(.00281)
Male	-.034	.014	-.0515	-.0804*	-.00283
	(.106)	(.013)	(.103)	(.0473)	(.0027)
High Tenure	.195	-.00645	.207*	.102**	-.0000749
	(.123)	(.0144)	(.122)	(.0469)	(.00338)
Crime Unit	.14*	-.0128	.153*	.0527	.000544
	(.0834)	(.00956)	(.0808)	(.0372)	(.00224)
Detective	-.0842	.00512	-.0903	-.0697**	.00284
	(.0866)	(.0111)	(.0803)	(.0336)	(.00221)
Early-Promoted to Sgt	.0965	.00541	.0922	.0408	.000939
	(.0857)	(.0106)	(.0796)	(.0319)	(.00195)
High Prior Arrests	-.0686	.012	-.0747	.035	.00846***
	(.1)	(.0117)	(.0927)	(.0389)	(.00255)
High Prior Public-Order Arrest Rate	.11	.00321	.105	.0301	.00124
	(.094)	(.0115)	(.0873)	(.0367)	(.00236)
High Racial Match with District Population	-.143*	-.0136	-.129*	-.0813***	-.00189
	(.0769)	(.00972)	(.0721)	(.0305)	(.00232)
Observations	430	430	430	430	430
Adj. R-squared	0.034	0.038	0.039	0.043	0.044
Mean Daily Arrests	3.53	.514	3.02	1.01	.0681

Note: This table regresses shrunken fixed effects (FEs) for LTs who supervised teams in the sample on their individual characteristics. On each day of actual supervision, we identify the LTs assigned to watches 1-3 and, when multiple LTs are present, prioritize the Watch Operations LT. After applying Bayesian shrinkage, we weight the regression by the number of days each LT supervised. All characteristics are entered in a regression together. Robust standard errors are reported. We zero out missings and include dummy variables to indicate missingness. All characteristics are binary indicators equal to one if the LT exceeds the sample median, except where the variable is inherently binary. See Appendix Section B.1 for details on variable construction.

Table A.5: Correlation between Relative LT FEs and Δ Characteristics

	Arrests FEs	Index FEs	Non-index FEs	Public Order FEs	Discretionary FEs
Black/Hispanic	-.163**	-.0165	-.145**	-.0761***	-.0129**
	(.0746)	(.0148)	(.0693)	(.0242)	(.0055)
Male	-.106*	-.0023	-.103*	-.0213	-1.4e-04
	(.0595)	(.0132)	(.0552)	(.0237)	(.0042)
Appointed Year	.0054	-2.2e-04	.0055	9.8e-04	9.1e-05
	(.0039)	(.0013)	(.0035)	(.002)	(4.5e-04)
Crime Unit	-.051	.0022	-.0535	-.0255	-.0083**
	(.0379)	(.0096)	(.038)	(.0182)	(.0041)
Detective	-.0276	.0037	-.0317	-.0317	-.0051
	(.0431)	(.0116)	(.0378)	(.0203)	(.0045)
Time to Sgt	-.0094	5.7e-04	-.01	-.0017	-.0012
	(.0074)	(.0015)	(.0069)	(.0034)	(9.4e-04)
Prior Arrests	-.0025	-2.8e-04	-.0022	-2.6e-04	-1.0e-05
	(.0016)	(2.1e-04)	(.0015)	(5.1e-04)	(6.6e-05)
Prior Public Order Rate	.294	.112*	.188	.341***	.0056
	(.238)	(.0585)	(.244)	(.0982)	(.0179)
Racial Match with District Population	-.0395	-.0015	-.0381	-.0196	-.0036
	(.0672)	(.0142)	(.0634)	(.0233)	(.0046)
Observations	265	265	265	265	265

Note: This table regresses the shrunken, relative LT FE for the 265 teams (district by watch by year) with exactly 2 predicted LTs on the difference in the characteristics between LT1 and LT2. Each characteristic is entered in a separate regression. Robust standard errors are reported. The full sample contains 265 lieutenant pairs for characteristics with complete coverage. Sample sizes are smaller for regressions using covariates with incomplete coverage, which require non-missing information for both lieutenants in the pair: time to lieutenant (224), time to sergeant (186), prior arrests (83), and prior arrest rates by severity (70). See Appendix Section B.1 for details on variable construction.

Table A.6: Correlation between Relative LT FEs and Δ Characteristics: All together

	(1) Arrests FEs	(2) Index FEs	(3) Non-index FEs	(4) Public Order FEs	(5) Discretionary FEs
Black/Hispanic	-0.122*	-0.016	-0.106*	-0.067***	-0.010*
	(0.066)	(0.015)	(0.062)	(0.024)	(0.006)
Male	-0.083	0.002	-0.086	-0.026	-0.004
	(0.059)	(0.014)	(0.057)	(0.027)	(0.005)
Appointed Year	0.004	-0.001	0.004	0.002	-0.000
	(0.004)	(0.001)	(0.004)	(0.002)	(0.001)
Crime Unit	-0.008	0.006	-0.015	-0.008	-0.004
	(0.041)	(0.010)	(0.041)	(0.020)	(0.005)
Detective	-0.027	0.002	-0.029	-0.028	-0.004
	(0.037)	(0.012)	(0.034)	(0.021)	(0.005)
Time to Sgt	-0.003	0.000	-0.003	0.001	-0.001
	(0.009)	(0.002)	(0.008)	(0.004)	(0.001)
Prior Arrests	-0.002	-0.000	-0.002	-0.000	-0.000
	(0.002)	(0.000)	(0.001)	(0.001)	(0.000)
Prior Public Order Arrest Rate	0.426	0.126**	0.306	0.348***	0.013
	(0.307)	(0.057)	(0.303)	(0.115)	(0.023)
Racial Match with District Population	-0.072	-0.001	-0.071	-0.027	-0.005
	(0.065)	(0.016)	(0.061)	(0.023)	(0.005)
Observations	265	265	265	265	265
Adj. R-squared	0.078	-0.006	0.079	0.019	0.033
Mean Daily Arrests	3.42	0.51	2.91	0.97	0.06

Note: This table regresses the shrunken, relative LT FE for the 265 teams (district by watch by year) with exactly 2 predicted LTs on the difference in the characteristics between LT1 and LT2. All characteristics are entered in a regression together. Robust standard errors are reported. We zero out missings and include dummy variables to indicate missingness. See Appendix Section B.1 for details on variable construction.

Table A.7: First Stage

	Any Black/Hisp. LT
Any Pred. Black/Hisp. LT	0.437***
	(0.027)
Observations	168,462
F-stat	258.87
Y mean	.192
Baseline FE and controls	Y

Notes: This table shows the results from the same regression as in Table 3 where the outcome is an indicator whether any Black of Hispanic LT actually worked in the given district-watch on the given day. *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table A.8: LT Presence and Substitution

	(1)	(2)
	No LT Present	Any New LT
Any Pred. Black/Hisp. LT	0.012	0.012
	(0.014)	(0.019)
Y mean	0.178	0.320
Observations	168,462	168,462

Notes: Each column reports estimates from a regression of the outcome on an indicator for a predicted Black or Hispanic LT, controlling for a predicted male LT and LT totals. Column (1) examines whether days with a predicted Black or Hispanic LT are more likely to have no LT present at all. Column (2) examines whether out-of-rotation substitution is more common on predicted Black or Hispanic LT days, where an out-of-rotation LT is defined as one who was not assigned to that district-watch in January of that year. All specifications include controls for any predicted male watch or field LT, and the predicted number of watch or field LTs, as well as district-watch-year fixed effects and calendar-date fixed effects. Standard errors are clustered at the team $\times$ year level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.9: IV Results: Arrests

	(1)	(2)	(3)	(4)
	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.289***	-0.234***	-0.023	-0.025
	(0.100)	(0.077)	(0.020)	(0.032)
Y mean	3.529	2.501	0.326	0.670
	PANEL B: ARREST TYPE			
	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.017	-0.272***	-0.164***	-0.017*
	(0.024)	(0.092)	(0.044)	(0.009)
Y mean	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y

Note: This table reports two-stage least squares (2SLS) estimates. We estimate the following system at the district-watch-year-by-day level:

$$\text{Any Black/Hisp. LT}_{jt} = \pi_1 \text{ Any Pred. Black/Hisp. LT}_{jt} + \pi_2 \text{ Any Pred. Male LT}_{jt} + \pi_3 \text{ Pred. Num. LTs}_{jt} + \alpha_j + \delta_t + u_{jt},$$

$$y_{jt} = \beta_1 \text{ Any } \widehat{\text{Black/Hisp.}} \text{ LT}_{jt} + \beta_2 \text{ Any Pred. Male LT}_{jt} + \beta_3 \text{ Pred. Num. LTs}_{jt} + \alpha_j + \delta_t + \epsilon_{jt}.$$

y_{jt} denotes a daily district-watch-level outcome (e.g. total arrests). *Any Pred. Black/Hisp. LT*$_{jt}$ is a dummy variable equal to one if any Black or Hispanic watch or field LT is predicted to be working on day t based on the predetermined day-off-group rotation schedule described in Section 3; this serves as the excluded instrument. α_j denotes district$\times$watch$\times$year fixed effects, so that all identifying comparisons are made within a district-watch-year. δ_t denotes calendar date fixed effects, which absorb common shocks to underlying crime rates. All specifications additionally control for whether any male watch or field LT is predicted to be working and for the predicted number of watch operations or field LTs. Panel A reports results for total arrests and for arrests of Black, white, and Hispanic individuals separately. Panel B disaggregates total arrests into the four arrest categories outlined in Section 3. Standard errors, clustered at the district-watch-year level, are in parentheses.

Table A.10: Arrests Linked to Court Conviction

	(1)	(2)	(3)	(4)
PANEL A: OVERALL GUILTY ARRESTS TO ALL ARRESTS				
	All			
Difference in Ratio	0.007			
	[0.303]			
Mean Ratio	0.567			
PANEL B: GUILTY ARRESTS RELATIVE TO ALL ARRESTS				
	Index	Non-index	Public Order	Discretionary
Difference in Ratio	0.007*	-0.001	-0.011*	-0.002
	[0.088]	[0.916]	[0.056]	[0.316]
Mean Ratio	0.108	0.459	0.155	0.015
PANEL C: GUILTY ARRESTS RELATIVE TO ALL ARRESTS WITHIN CATEGORY				
	Index	Non-index	Public Order	Discretionary
Difference in Ratio	-0.002	0.005	0.005	0.044
	[0.902]	[0.447]	[0.770]	[0.501]
Mean Ratio	0.772	0.533	0.524	0.759
Observations	84,231	84,231	84,231	84,231
Baseline FE and controls	Y	Y	Y	Y

Note: We define an arrest as leading to a conviction if the case is disposed of as either a misdemeanor or a felony. Otherwise, an arrest is defined as non-guilty (if the case is dismissed, not matched to the court data, or does not have clear case outcomes). We estimate coefficients in regressions where the dependent variables are guilty arrests and total arrests. We add the estimated coefficients for guilty and overall arrests to their respective means and compare the resulting ratio with the ratio of the means. The p-values are in brackets. In Panel A, we present results for the difference in the ratios of overall guilty arrests to all arrests. In Panel B, we present the difference in the ratios of guilty arrests of specific arrest types relative to overall arrests. Panel C presents the difference in the ratios of guilty arrests of a given type to all arrests of the same type.

Table A.11: Effects of Sole Black or Hispanic Lieutenant Leadership

	Arrests by Race of Arrestee				Arrests by Type			
	All	Black	White	Hispanic	Serious	Non-Index	Public Order	Discretionary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Pred. BH LT (Sole Leadership)	-0.102**	-0.087**	-0.014	0.000	-0.003	-0.099**	-0.057***	-0.012***
	(0.051)	(0.044)	(0.009)	(0.015)	(0.011)	(0.048)	(0.020)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462

Notes: This table reports estimates restricting to days on which all LTs present are Black or Hispanic. The outcome variables are daily team-level arrest counts. The indicator “Sole Leadership” equals one on days when the predicted LT assignment is exclusively Black or Hispanic — that is, no white LT is predicted to be present on that team-day. All specifications include the baseline fixed effects and controls used in Table 3. Standard errors are clustered at the district-watch level.

Table A.12: Share Black/Hisp LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: SHARE BLACK/HISP LT				
Share Black/Hisp. LT	-0.169***	-0.135***	-0.018*	-0.013	-0.008	-0.160***	-0.096***	-0.012**
	(0.059)	(0.049)	(0.011)	(0.017)	(0.013)	(0.055)	(0.024)	(0.005)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168462	168462	168462	168462	168462	168462	168462	168462
				PANEL B: DUMMIED OUT SHARE BLACK/HISP LT				
$0 <$ Share Black/Hisp. LT ≤ 0.5	-0.101**	-0.083***	-0.003	-0.012	-0.008	-0.093**	-0.055***	-0.003
	(0.043)	(0.032)	(0.009)	(0.016)	(0.012)	(0.038)	(0.020)	(0.004)
$0.5 <$ Share Black/Hisp. LT ≤ 1	-0.164***	-0.131***	-0.019*	-0.011	-0.008	-0.156***	-0.092***	-0.012**
	(0.060)	(0.050)	(0.011)	(0.017)	(0.013)	(0.056)	(0.024)	(0.005)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: In this table, we replace the indicator for any predicted Black or Hispanic LT with the share of Black or Hispanic LTs among predicted watch and field LTs. In 98% of cases for which there is a Black or Hispanic LT, this share is either 0.5 or 1 (with about an even split). In very few cases it is $1/3$ or $2/3$. In Panel B, we investigate effects on dummies for predicted share Black or Hispanic LT being greater than 0 but less than or equal to $1/2$ (effectively $1/2$) and for the predicted share being greater than $1/2$ (effectively 1), to more flexibly examine how the share of Black or Hispanic LTs affects outcomes. All else is as in Table 3, except that here we also flexibly control for the number of predicted LTs using a full suite of dummies as opposed to doing so linearly, given its direct implications for the shares.

Table A.13: Disaggregating LT Roles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. Watch Op. LT	-0.163***	-0.114**	-0.029***	-0.015	-0.000	-0.163***	-0.095***	-0.010*
	(0.061)	(0.047)	(0.010)	(0.019)	(0.013)	(0.056)	(0.027)	(0.006)
Any Pred. Black/Hisp. Field LT	-0.031	-0.040	0.018	-0.010	-0.019	-0.012	-0.027	0.001
	(0.053)	(0.046)	(0.012)	(0.014)	(0.014)	(0.050)	(0.024)	(0.005)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168462	168462	168462	168462	168462	168462	168462	168462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
P-value for Watch LT = Field LT	0.095	0.274	0.002	0.804	0.287	0.050	0.062	0.162

Note: Here we allow for separate effects of Any Pred. Black/Hispanic watch operations LT and Any Pred. Black/Hispanic field LT, as opposed to pooling these two groups. We define an LT as a predicted watch operations LT if he or she worked as a WOL more than 75% of the work days in January, and as a field LT if he or she worked as a WOL less than 75% of the days. Instead of controlling for the gender and predicted number of LTs overall, we also now separately control for the predicted number of field LTs, the predicted number of watch operations LTs, any predicted male field LT, and any predicted male watch operations LT. P-values for a test of equality of the two coefficients are also reported.

Table A.14: Pooled vs. Disaggregated Race of Lieutenant: Summary Across Three Specifications

	(1) All	(2) Black	(3) White	(4) Hispanic	(5) Serious	(6) Non-Index	(7) Public Order	(8) Discretionary
PANEL A: BASELINE REDUCED FORM								
Any Pred. Black/Hisp. LT	−0.126***	−0.102***	−0.010	−0.011	−0.007	−0.119***	−0.072***	−0.007*
	(0.043)	(0.033)	(0.008)	(0.014)	(0.011)	(0.040)	(0.019)	(0.004)
Any Pred. Black LT (no controls)	−0.074	−0.057	−0.007	−0.009	−0.016	−0.057	−0.039	−0.004
	(0.082)	(0.071)	(0.010)	(0.017)	(0.016)	(0.078)	(0.029)	(0.006)
Any Pred. Hispanic LT (no controls)	−0.127***	−0.100***	−0.010	−0.013	−0.002	−0.125***	−0.081***	−0.007
	(0.048)	(0.036)	(0.011)	(0.017)	(0.012)	(0.045)	(0.023)	(0.005)
P-value: Black = Hispanic	0.575	0.588	0.812	0.867	0.441	0.456	0.249	0.629
Any Predicted Black LT (controls)	-0.100	-0.075	-0.009	-0.015	-0.017	-0.083	-0.053*	-0.004
	(0.077)	(0.067)	(0.010)	(0.017)	(0.017)	(0.072)	(0.028)	(0.006)
Any Predicted Hispanic LT (controls)	-0.114**	-0.091***	-0.009	-0.010	-0.003	-0.111***	-0.079***	-0.007
	(0.045)	(0.034)	(0.010)	(0.016)	(0.011)	(0.042)	(0.022)	(0.005)
P-value for Black LT = Hisp LT	0.865	0.827	0.979	0.849	0.477	0.732	0.451	0.657
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
PANEL B: CORRELATION BETWEEN RELATIVE LT FIXED EFFECTS AND Δ LT CHARACTERISTICS								
Black/Hispanic	-0.122*	-0.116**	-0.007	0.006	-0.016	-0.106*	-0.067***	-0.010*
	(0.066)	(0.054)	(0.015)	(0.021)	(0.015)	(0.062)	(0.024)	(0.006)
Black	-0.201	-0.137	-0.022	-0.038	-0.033	-0.170	-0.090	0.001
	(0.163)	(0.119)	(0.033)	(0.059)	(0.036)	(0.154)	(0.057)	(0.009)
Hispanic	-0.085	-0.106**	0.000	0.027	-0.007	-0.076	-0.057*	-0.015**
	(0.066)	(0.051)	(0.015)	(0.028)	(0.016)	(0.064)	(0.029)	(0.007)
Observations	265	265	265	265	265	265	265	265
P-value: Black = Hisp	0.524	0.805	0.545	0.398	0.538	0.592	0.636	0.189
PANEL C: EFFECTS OF SOLE BLACK OR HISPANIC LIEUTENANT LEADERSHIP								
Any Pred. BH LT (Sole Leadership)	−0.102**	−0.087**	−0.014	0.000	−0.003	−0.099**	−0.057***	−0.012***
	(0.051)	(0.044)	(0.009)	(0.015)	(0.011)	(0.048)	(0.020)	(0.004)
Any Black LT	−0.136	−0.126	−0.012	0.002	−0.022	−0.114	−0.063**	−0.015**
	(0.103)	(0.086)	(0.014)	(0.027)	(0.020)	(0.096)	(0.031)	(0.006)
Any Hispanic LT	−0.111*	−0.094*	−0.016	0.002	0.004	−0.115**	−0.058**	−0.012**
	(0.060)	(0.053)	(0.011)	(0.018)	(0.012)	(0.057)	(0.027)	(0.005)
P-value: Black = Hispanic	0.836	0.756	0.802	0.988	0.253	0.993	0.909	0.727
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068

Notes: Each panel shows results first for Black and Hispanic LTs pooled, then disaggregated into Black and Hispanic separately. Panel A shows our baseline reduced-form estimates change when the race of the LT is disaggregated. The first row repeats the baseline effects of having any predicted Black or Hispanic LT. Next, we report the effects using separate indicators for any predicted Black and any predicted Hispanic LT, without additional controls (no controls). Third, we add controls for team composition – i.e. the outcome variables from Table 2 (controls). Panel B shows how the results in Table A.6, which estimate the relationship between the shrunken relative LT fixed effects and LT characteristics, change when the race of the LT is disaggregated. Panel C shows how the results from sole leadership specification in Table A.11, which focuses on days when only Black or Hispanic LTs are present, change when the race of the LT is disaggregated.

Table A.15: Controlling for Predicted and Actual Officer and Sergeant Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: CONTROLLING FOR PREDICTED TEAM COMPOSITION				
Any Pred. Black/Hisp. LT	-0.128***	-0.104***	-0.010	-0.011	-0.009	-0.119***	-0.075***	-0.007*
	(0.041)	(0.032)	(0.008)	(0.014)	(0.010)	(0.038)	(0.018)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168462	168462	168462	168462	168462	168462	168462	168462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
				PANEL B: CONTROLLING FOR ACTUAL TEAM COMPOSITION				
Any Pred. Black/Hisp. LT	-0.120***	-0.097***	-0.009	-0.011	-0.008	-0.112***	-0.070***	-0.007*
	(0.041)	(0.032)	(0.008)	(0.014)	(0.010)	(0.037)	(0.018)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: We report the effect of Any Pred. Black/Hisp LT on the same outcomes as in Table 3, controlling first in Panel A for all the predicted team composition variables in Table 2 panels A-C. In Panel B, we instead control for the (potentially endogenous) characteristics of the actual workforce (i.e. all the outcome variables in panels A-C of Table A.16).

Table A.16: Actual Team Composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share Black	Share Hisp.	Share White	Share Male	Pred. Num.	Prev. Arrests	Prev. Complaints	Prev. Force
				Panel A: Actual Officer + Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0012	0.0048**	-0.0042	0.0017	-0.0445	-0.5164	-0.0027	-0.0033
	(0.0021)	(0.0021)	(0.0029)	(0.0023)	(0.2040)	(0.3150)	(0.0038)	(0.0066)
Y mean	0.191	0.167	0.476	0.756	51.358	35.488	0.181	0.432
				Panel B: Actual Officer Characteristics				
Any Pred. Black/Hisp. LT	-0.0007	0.0055**	-0.0027	0.0031	-0.0478	-0.5890*	-0.0035	-0.0072
	(0.0023)	(0.0021)	(0.0031)	(0.0026)	(0.1892)	(0.3277)	(0.0040)	(0.0069)
Y mean	0.201	0.168	0.454	0.749	46.567	37.231	0.177	0.442
				Panel C: Actual Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0186***	-0.0015	-0.0160*	-0.0105*	0.0098	0.4765	0.0032	0.0355**
	(0.0054)	(0.0070)	(0.0083)	(0.0063)	(0.0339)	(0.4805)	(0.0107)	(0.0152)
Y mean	0.096	0.160	0.687	0.823	4.792	18.529	0.224	0.333
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table examines whether district-watch-days with predicted minority LTs have different actual team compositions. The specification is as in Equation 3 replacing the outcome variable with the stated variable. We set the few missing outcome values that arise equal to 0 and include an extra dummy variable control for missing values. Panels A pertains to all officers and sergeants. Panels B pertains to just officers. Panels C pertains to just sergeants. Columns (6)-(8) examine the total yearly arrests, complaints, and use of force from two years prior to the current year. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A.17: Independent Effects of Minority Sergeants

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
Any Pred. Black/Hisp. LT	-0.124***	-0.101***	-0.010	-0.010	-0.008	-0.116***	-0.072***	-0.007*
	(0.042)	(0.033)	(0.008)	(0.014)	(0.010)	(0.039)	(0.019)	(0.004)
Share Black/Hisp Sergeants	-0.113*	-0.061	-0.016	-0.035	-0.046***	-0.067	0.014	-0.021***
	(0.065)	(0.055)	(0.014)	(0.022)	(0.018)	(0.060)	(0.035)	(0.007)
Share Male Sergeants	-0.074	-0.063	-0.038**	0.019	-0.079***	0.005	-0.040	-0.012
	(0.077)	(0.061)	(0.018)	(0.027)	(0.023)	(0.074)	(0.046)	(0.010)
Number of Predicted Sergeants	0.028***	0.021**	0.001	0.007*	-0.001	0.030***	0.004	0.000
	(0.011)	(0.009)	(0.002)	(0.004)	(0.003)	(0.010)	(0.006)	(0.001)
Prev. Arrests Sergeants	0.005***	0.004***	0.001***	0.001	0.000	0.005***	0.001*	0.000
	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Prev. Complaints Sergeants	-0.007	-0.003	-0.009	0.005	-0.012	0.005	-0.019	0.006
	(0.060)	(0.049)	(0.011)	(0.020)	(0.016)	(0.057)	(0.028)	(0.007)
Prev. Use of Force Sergeants	-0.062	-0.032	-0.014	-0.015	-0.006	-0.055	-0.011	0.001
	(0.048)	(0.039)	(0.010)	(0.015)	(0.011)	(0.045)	(0.023)	(0.005)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to Table 3, but additionally controls for predicted sergeant composition and characteristics. In addition to the baseline LT controls, specifications include the predicted share of Black or Hispanic sergeants, predicted share of male sergeants, predicted number of sergeants, predicted average prior arrest activity of sergeants, predicted average prior complaint history of sergeants, and predicted average prior force-complaint history of sergeants. All specifications include the baseline fixed effects and controls from Table 3. Standard errors are clustered at the district-watch-year level. The mean share of minority sergeants is 0.26 with a standard deviation of 0.21 and the mean number of sergeants is 5.5.

Table A.18: Individual-Level Arrests

	(1)	(2)	(3)	(4)
	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-.00139*	-.00103*	-.000061	-.000289
	(.000726)	(.000541)	(.000142)	(.000264)
Y mean	.0481	.0329	.0048	.00994
FE	Baseline	Baseline	Baseline	Baseline
Any Pred. Black/Hisp. LT	-.0012**	-.00103**	.0000264	-.000209
	(.000543)	(.000421)	(.000126)	(.000216)
Y mean	.0481	.0329	.0048	.00994
FE	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer
	PANEL B: ARREST TYPE			
	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	.00011	-.0015**	-.000762**	-.0000171
	(.000157)	(.000685)	(.000307)	(.0000542)
Y mean	.00756	.0405	.0138	.000971
FE	Baseline	Baseline	Baseline	Baseline
Any Pred. Black/Hisp. LT	.000111	-.00131**	-.000777***	-.0000247
	(.000148)	(.000511)	(.000262)	(.0000512)
Y mean	.00756	.0405	.0138	.000971
FE	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer

Note: In this table, we report results from an alternative specification where observations are at the individual officer-day level. We restrict attention to officers who hold the titles of "Police Officer" or "Sergeant of Police." To avoid double counting, we only count arrests for which the officer was the first arresting officer. We regress the daily individual outcomes of officers who are actually present in the team on that day on a binary variable of whether a Black or Hispanic LT is predicted to be present. In Panel A, the outcome variables are arrests as well as arrests of different civilian race. In the sub-panel at the bottom of Panel A, we include in the regression individual officer FEs. We repeat a similar comparison in Panel B where the outcome variables are different types of arrests.

Table A.19: Heterogeneity by Prior Exposure to Black or Hispanic Lieutenants

	(1) All	(2) Index	(3) Non-Index	(4) Public Order	(5) Discretionary
Any Pred. Black/Hisp. LT	-.00728**	-.000385	-.00689**	-.00177	-.000186
	(.0034)	(.00092)	(.0029)	(.00132)	(.000332)
Any Pred. Black/Hisp. LT × Prior Exposure	.00614*	.000578	.00556*	.000993	.000194
	(.00343)	(.000923)	(.00291)	(.00135)	(.000337)
Y mean	.0464	.00715	.0392	.0134	.000916
Observations	6523585	6523585	6523585	6523585	6523585
FE	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer
Any Pred. Black/Hisp. LT	-.0102***	-.00164	-.00852***	-.00203	-.000072
	(.0038)	(.00104)	(.00328)	(.00151)	(.000314)
Any Pred. Black/Hisp. LT × Prior Exposure	.00563*	.000418	.00521*	.000791	.000163
	(.00339)	(.000948)	(.00285)	(.00128)	(.000328)
Y mean	.0462	.00715	.039	.0133	.000912
Observations	6432749	6432749	6432749	6432749	6432749
FE	+ Char. Interactions	+ Char. Interactions	+ Char. Interactions	+ Char. Interactions	+ Char. Interactions

Notes: This table examines how the effects of LT assignment vary with officers' prior exposure to Black or Hispanic LTs. The sample includes police officers and sergeants. The outcome variables are daily arrests per officer. Each column reports estimates from separate regressions that include indicators for predicted Black/Hispanic LT assignment, prior exposure to Black/Hispanic LTs, and their interactions. We report the interaction between assignment to a Black/Hispanic LT and prior exposure. The omitted group corresponds to officers with no prior exposure to Black/Hispanic LTs. Prior exposure is measured based on officer history over the previous three years. Officers with no prior years in the sample are excluded. All specifications include baseline fixed effects (district-watch-year and day-of-week) and officer fixed effects. Standard errors are clustered at the district-watch level. The lower panel augments by additionally controlling for interactions between predicted Black/Hispanic LT assignment and four officer characteristics: an indicator for Black or Hispanic officer race, an indicator for above-median tenure, an indicator for having remained in the same district-watch for at least some portion of the prior three years, and an indicator for above-median average minority district share over the prior three years. * p<0.10, ** p<0.05, *** p<0.01.

Table A.20: Heterogeneous Effects by Neighborhood Characteristics

	(1) All	(2) Index	(3) Non-Index	(4) Public Order	(5) Discretionary
Panel A: Interaction with Above-Median District % Black or Hispanic					
Any Pred. Black/Hisp. LT	-0.067*	-0.009	-0.058*	-0.039*	-0.001
	(0.037)	(0.015)	(0.034)	(0.020)	(0.003)
Pred. Black/Hisp. LT × High % Black/Hisp.	-0.102	0.003	-0.104	-0.055*	-0.010
	(0.073)	(0.020)	(0.067)	(0.033)	(0.007)
Y mean	3.529	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462
Panel B: Interaction with Above-Median Pre-Period (2012) Index Crime					
Any Pred. Black/Hisp. LT	-0.077	-0.005	-0.071	-0.053	-0.003
	(0.069)	(0.015)	(0.064)	(0.034)	(0.004)
Pred. Black/Hisp. LT × High Pre-Period Crime	-0.081	-0.003	-0.078	-0.031	-0.006
	(0.083)	(0.020)	(0.077)	(0.039)	(0.007)
Y mean	3.529	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462
Panel C: Interaction with Above-Median Pre-Period (2012) Arrests					
Any Pred. Black/Hisp. LT	-0.114*	-0.013	-0.101	-0.036	-0.002
	(0.067)	(0.014)	(0.061)	(0.028)	(0.004)
Pred. Black/Hisp. LT × High Pre-Period Arrests	-0.018	0.008	-0.026	-0.051	-0.007
	(0.082)	(0.019)	(0.076)	(0.035)	(0.007)
Y mean	3.529	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462

Notes: Each column corresponds to an arrest outcome; each panel introduces a different interaction with a neighborhood characteristic. In all panels the baseline coefficient "Any Pred. Black/Hisp. LT" gives the effect in the below-median group, and the interaction term gives the differential effect in the above-median group. Panel A interacts the predicted minority LT indicator with an indicator for above-median combined Black and Hispanic share of district residents. Panel B interacts with an indicator for above-median district-level 2012 index crimes per capita. Panel C interacts with an indicator for above-median district-level 2012 arrests per capita. Above-median splits are defined within the sample of 21 CPD districts. All specifications include the baseline fixed effects and controls used in Table 3. Standard errors are clustered at the team × year level in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.21: Restricting to District-watch-Years with Variation in Any Predicted Black/Hispanic LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.129**	-0.114**	-0.008	-0.005	-0.005	-0.124**	-0.067***	-0.006
	(0.059)	(0.046)	(0.009)	(0.018)	(0.012)	(0.054)	(0.023)	(0.004)
Y mean	3.820	2.928	0.291	0.572	0.542	3.279	1.079	0.074
Observations	42783	42783	42783	42783	42783	42783	42783	42783
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to the one in Table 3 except it restricts to district-watch-years with variation in the dummy for having any predicted Black or Hispanic LT. That is, it drops teams with only non-Black and Hispanic predicted LTs or (the few) teams that always have a Black or Hispanic LT predicted on all days, which is possible with 2 or more LTs.

Table A.22: White LTs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. White LT	0.118*	0.110*	0.010	-0.006	0.004	0.114*	0.062***	0.014***
	(0.068)	(0.061)	(0.011)	(0.017)	(0.015)	(0.063)	(0.022)	(0.005)
Y mean	3.820	2.928	0.291	0.572	0.542	3.279	1.079	0.074
Observations	42783	42783	42783	42783	42783	42783	42783	42783
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to the one in Table A.21 except it replaces the indicator for any Black or Hispanic LT with an indicator for any white operations or field LT predicted on duty.

Table A.23: Alternate Inference

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				Panel A: Baseline				
Any Pred. Black/Hisp. LT	-0.126***	-0.102***	-0.010	-0.011	-0.007	-0.119***	-0.072***	-0.007*
	(0.043)	(0.033)	(0.008)	(0.014)	(0.011)	(0.040)	(0.019)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
				Panel B: Clustering at district-watch level				
Any Pred. Black/Hisp. LT	-0.126**	-0.102***	-0.010	-0.011	-0.007	-0.119**	-0.072***	-0.007*
	(0.052)	(0.035)	(0.008)	(0.017)	(0.011)	(0.046)	(0.020)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
				Panel C: Randomization Inference				
Any Pred. Black/Hisp. LT	-0.125***	-0.094***	-0.009	-0.019	-0.008	-0.117***	-0.074***	-0.007*
Rand. inference se	(0.042)	(0.035)	(0.007)	(0.012)	(0.009)	(0.040)	(0.018)	(0.004)
Rand. inference p-value	[0.0015]	[0.0060]	[0.2150]	[0.1180]	[0.3645]	[0.0025]	[0.0000]	[0.0630]
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168462	168462	168462	168462	168462	168462	168462	168462

Note: This Table reports the same estimates as in Table 3 using different inference procedures. Panel A is as in Table 3. Panel B clusters standard errors at the district-watch level instead of the district-watch-year level. Panel C performs a randomization inference procedure in which teams that have variation in any predicted Black or Hispanic LT have that indicator randomly reassigned, based on randomly choosing one of the 6 day off groups. We employ 2,000 trials; the resulting distributions of estimates (shown in Figure A.5) appear approximately normal and are centered around 0. The reported standard errors in panel C correspond to the standard deviation of these 2,000 estimates, and the p-values correspond to the share of these 2,000 trials that produce estimates larger in absolute value than our baseline.

Table A.24: Poisson Maximum Likelihood Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.030***	-0.031***	-0.029	-0.018	-0.013	-0.033***	-0.063***	-0.103**
	(0.010)	(0.010)	(0.028)	(0.023)	(0.019)	(0.011)	(0.017)	(0.051)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.073
Observations	168462	168462	168399	168462	168462	168462	168462	157505
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
Effect of Any Pred. Black/Hisp LT	-0.104	-0.078	-0.009	-0.012	-0.007	-0.098	-0.063	-0.007
p-value	0.003	0.002	0.289	0.430	0.494	0.002	0.000	0.038

Note: This table uses Poisson pseudo-maximum likelihood estimation (PPML) instead of OLS to estimate the effect of any predicted Black or Hispanic LT. All else is as in Table 3. For ease of comparability the implied marginal effect of going from 0 to any predicted Black or Hispanic LT and its corresponding p-value are reported at the bottom of the table.

Table A.25: Stops in Various Neighborhoods

	(1)	(2)	(3)	(4)
	Stops	High Black	High White	High Hispanic
Any Pred. Black/Hisp. LT	-0.1120	-0.0696	-0.0375	-0.0222
	(0.1175)	(0.0559)	(0.0388)	(0.0412)
Observations	168,462	168,462	168,462	168,462
Y mean	11.4984	3.6958	2.3005	3.0928
Baseline FE and controls	Y	Y	Y	Y

Note: This table runs the same regression as in Table 3 for 4 additional outcomes: total stops, total stops in high-Black, high-white, and high-Hispanic beats. A "High-Black" beat is defined on a district-by-district basis and equals 1 for all beats in the district with a population share Black exceeding the 75th percentile of share Black for that district. High-white and high-Hispanic beats are defined analogously.

B Additional Data and Variable Construction Details

B.1 Construction of Lieutenant Characteristics

This appendix section describes the construction of the lieutenant-level characteristics used throughout the paper.

Promotion and tenure histories. We begin with promotion records obtained through FOIA requests to CPD, which contain appointment dates and promotion years for officers promoted to sergeant and lieutenant from 2004. We supplement these records using assignment histories from OpenOversight and manual LinkedIn searches in cases where promotion dates are missing from the administrative files. OpenOversight supplementation increases coverage of promotion-to-sergeant dates by 286 lieutenants and promotion-to-lieutenant dates by 186 lieutenants, with additional smaller gains from LinkedIn searches (8 for sergeants and 6 for lieutenants). Using these records, we construct years from appointment to promotion to sergeant and lieutenant, along with indicators for early promotion defined as promotion times below the sample median. Tenure is measured as years since appointment as of 2012.

Prior arrest histories. To measure prior enforcement activity, we use arrest records from the period before lieutenants became supervisors in our sample. We first construct a lieutenant-by-year panel from appointment (or 1999, whichever comes later) through 2011. We then restrict attention to years before promotion to sergeant and before 2012, so that arrest histories reflect activity while serving as patrol officers rather than supervisors.

We construct a measure of prior arrest activity by estimating a regression of annual arrest counts on tenure-year indicators and lieutenant fixed effects over this pre-supervisory officer-year panel. High prior arrests equals one if the lieutenant fixed effect from this regression exceeds the sample median. We also construct prior arrest-rate measures, defined as the share of patrol-officer arrests belonging to a given category (e.g., public-order, index, non-index, or discretionary arrests). These arrest-rate measures are calculated only among lieutenants with at least 30 prior arrests.

Prior assignments. We use historical unit assignment records to identify whether lieutenants previously served in detective units or specialized crime-enforcement units before 2012. Detective is an indicator for prior assignment to detective units, while crime unit indicates prior service in specialized enforcement units, including tactical and narcotics-related assignments.

Complaints and awards. We merge complaint records from CPD complaint databases to construct measures of citizen complaints received before 2012. We separately identify complaints related to arrests, force, and operations/personnel issues using complaint-category codes. We also merge historical awards records and construct measures of total awards and meritorious awards received before 2012.

Racial concordance measures. To measure racial concordance between lieutenants and the districts they supervise, we first define a lieutenant-day indicator equal to one if the lieutenant's race matches the largest racial group in the district population. District 17 has roughly comparable white and Hispanic population shares, so both white and Hispanic LTs are treated as racially matched in that district. We then average this indicator across all supervision days for each lieutenant and define a high racial-match indicator equal to one if this average exceeds the sample median.

Binary indicators. Unless otherwise noted, lieutenant characteristics used in the main analysis are converted into binary indicators equal to one if the underlying measure exceeds the sample median.

B.2 Construction of OEMC Call Outcomes

This appendix section describes the construction of call-type and dispatch-time outcomes using OEMC Police Computer Aided Dispatch records.

We classify dispatch descriptions into broad call categories using the final 911 dispatch description. We first convert all dispatch descriptions to uppercase. We then classify calls as *violent* if the description contains terms such as battery, assault, abduction, kidnapping, robbery, sex offense, gun, knife, bomb, shots, or stabbing. We classify calls as *property* if the description contains terms such as burglary, auto, property, criminal damage, trespass, theft, arson, forgery, deceptive practice, or looting. We classify calls as *quality of life* if the description contains terms such as disturbance, loitering, gambling, fireworks, curfew violation, animal complaints, prostitution, narcotics, or vice complaints. Finally, we classify calls as *community assistance* if the description contains terms such as assist citizen or community.

Our analysis focuses on three dispatch activity categories: i) we define *enforcement activity* as calls classified as violent or property-related, ii) we define *social-order activity* as calls classified as quality-of-life related, and iii) we define *community activity* as calls classified as community assistance. For each category, we construct counts of calls in that category occurring within each district-watch-day.

We define time to dispatch as the number of minutes between the event time and dispatch time. Because dispatch timestamps are consistently available beginning in 2014, these measures are constructed for 2014 onward. We exclude non-positive dispatch times and calls whose descriptions indicate non-dispatched or administrative activity, including missions, details, lunch, overtime, and on-view events. We winsorize time to dispatch at the 99th percentile. We also construct priority-specific dispatch-time measures separately for calls coded as priority 1, priority 2, and priority 3.