Do Teams Perform Differently Under Black and Hispanic Leaders? Evidence from the Chicago Police Department*

Matthew Gudgeon, Andrew Jordan, and Taeho Kim

December 2023

Abstract

One oft-suggested policy to improve equity in policing involves increasing racial diversity among police leadership, but there is little evidence on how this would affect outcomes. We show, in the context of the Chicago Police Department (CPD), that police teams managed by Black and Hispanic supervisors make fewer low-level arrests. To overcome potentially endogenous absences, we exploit the pre-determined, rotational nature of the CPD operations calendar and compare within unit-watch outcomes on days when Black and Hispanic police lieutenants (LTs) are expected on duty to (otherwise similar) days when they have a day off and another LT is in charge.

*All opinions expressed in this manuscript are those of the authors. We thank Oeindrila Dube, Greg Distelhorst, Ben Pyle, Roman Rivera, CarlyWill Sloan, Anna Weber, and seminar participants at APPAM, the University of Toronto, and West Point. All errors remain ours. Gudgeon: Tufts University (matthew.gudgeon@tufts.edu); Jordan: Washington University in St Louis (awjordan@wustl.edu); Kim: University of Toronto (thk.kim@utoronto.ca).

1 Introduction

In the aftermath of recent police shootings, the U.S. has seen growing public outrage over police brutality incidents involving minority citizens. These events represent the latest eruption of long-standing concerns about over-policing and disparate treatment in minority communities. There is also increasing recognition of the costs of such policing, for public trust and legitimacy, for minorities, and for police departments themselves (Bor et al., 2018; Ang, 2021; Owens and Ba, 2021; Alexander et al., 2022; Ang et al., 2023). A commonly suggested policy response is to increase racial diversity and representation in police departments (US Department of Justice, 2016). Indeed, police department have long employed a strategy of pivoting to minority leadership to address racial turmoil.[1]

While an abundance of evidence has shown that minority officers engage in less aggressive policing practices than their white peers (Ba et al., 2021b; Goncalves and Mello, 2021; Hoekstra and Sloan, 2022; Rivera, 2022), the impact of minority police leaders, who may not be directly involved in fieldwork, remains less clear.[2] This lack of evidence is not surprising given the fundamental data and identification challenges involved with estimating plausibly causal effects of leaders. On the one hand, minority leadership in police departments could wield significant influence. They could influence crime strategies, enforce standards, mentor officers, intervene with those officers in need of intervention, and affect culture. If so, interventions targeting leaders could prove highly cost-effective. On the other hand, it is not always clear how much scope and willingness police supervisors have to oversee their subordinates or influence department culture, nor how accountable they are for their management practices (US Department of Justice, 2017).

In this paper, we investigate the effects of high-ranking Black and Hispanic police officers (lieutenants or LTs) on daily shift (district by watch) level outcomes in the context of the Chicago Police Department (CPD). We use quasi-experimental variation in work schedules generated by the CPD's operations calendar. While each district has a single commander and captain jointly in charge of its crime suppression activities, LTs are the first-line supervisors in charge of the more routine functioning of operations. There must always be a watch operations LT on duty on each *watch*. This stands in contrast to the district commander who cannot work 24 hours a day, 7 days a week. Watch operations LTs, stationed in the district headquarters, oversee both personnel/staffing

[1] For example, Willie L Williams was appointed the first Black police chief in 1992 at the Los Angeles Police Department after riots erupted over the beating of Rodney King; in 2016, two years after Michael Brown was shot to death by a police officer in Ferguson, MO, Delrish Moss became the first Black police chief in that city (Capital B News, 2022). A recent article quotes Anthony Chapa, the executive director of the Hispanic American Police Command Officers Association: "the best way for the public to regain trust with law enforcement is to have more minority officers in departments, especially in high-ranking positions" (ABC News, 2020).

[2] Bulman (2019), who studies how the race of U.S. sheriffs affects black-white arrest ratios, is an important exception.

and review and approve all arrest reports and complaints, among others duties outlined in the CPD directives (see Figure 1). For this reason, watch operations LTs have traditionally been referred to as watch commanders. LTs can also serve in the role of field LTs, in charge of responding to and assuming command of major field incidents within their respective districts. Field LTs can and regularly do assume the role of watch operations LT (such as on a day off) and vice-versa. As a result, our baseline results examine the effect of having a Black and Hispanic field *or* watch operations LT on team outcomes.[3]

We cannot isolate the effect of Black and Hispanic LTs using simple comparisons across police districts since LTs of different races may work in environments with different crime conditions. Even within the same district and watch, we cannot use the presence of a Black or Hispanic LT on a particular day because this too is potentially endogenous. LTs may transfer out, take vacations, or be put on leave, and these actions could be related to team-level performance or its determinants (e.g. crime conditions) and correlated with LT race. Indeed, among weekend days when LTs are expected on duty, Black and Hispanic LTs are about 1pp more likely than white LTs to take leave (such as a vacation). This indicates that there is some reason to believe LT choices of when to work could be correlated with both underlying crime patterns and race.[4]

Instead, we exploit a quasi-experiment generated by the 4-day on, 2-day off rotation schedule of LTs (and their subordinates). The CPD's 22 districts follow an operations calendar that assigns officers to one of six day-off groups (DOGs). On any given calendar day, two DOGs are off-duty, while the other four are on duty. The day-off group schedule for the year is predetermined in advance. Each shift — called a watch — in a district is managed by 1-3 lieutenants, who also rotate days worked based on their DOGs. To fix ideas, take a district-watch with one white and one Black LT in different day-off groups (as would always be the case). The DOG schedule creates a set of days in which the Black LT is supposed to be on duty, on some of which they will have sole command and on others of which both LTs will be on duty, and a set of days in which the Black LT has pre-scheduled days off and the white LT assumes sole command. Our empirical approach examines how total arrests at the daily, district-by-watch level, differ on days when the Black or

[3]LTs can also serve in a third role, that of tactical LTs. These LTs oversee tactical teams and do not commonly fill in for the field or watch operations LTs. They work on a separate schedule that does not directly correspond to the normal 3 watches. As a result, examining their role, which we also do, requires a different empirical strategy and necessitates looking at district-level (as opposed to district-watch) level outcomes.

[4]We note that this concern may be more salient at higher ranks, since lower ranks may find it harder to get approved leave. 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 lieutenant called in sick. He said he had hives. ... The following roll call, on New Year's Day, the lieutenant returned in high spirits and with no sign of hives" (Moskos, 2008).

Hispanic LT is predicted to be on duty relative to days which they are scheduled to have off.[5]

The predetermined and rotational nature of these work schedules ensures that leadership assignments within a team are not systematically correlated with underlying crime patterns. Consistent with this, days predicted to have Black or Hispanic LTs do not differ, on average, in their underlying crime conditions, relative to days these leaders are scheduled to have off. Daily arrests in other police districts (both overall and neighboring) in the same watch, 911 calls in both the given district-watch and in neighboring district-watches, and predicted arrests in the given district-watch are all uncorrelated with whether a Black or Hispanic LT is predicted to be on duty.

Since subordinate officers share the same set of possible day-off groups as LTs, it is at least possible that there is some same-race or other sorting into days off, in turn engendering differences in team composition across our treatment and control groups. Some officers will work with a given LT on all four days that they work (in every 6 day cycle), while others will work with the given LT on 2 or 3 days (in every cycle). As a result, while all officers work with all LTs and vice versa, some do so more frequently than others. Nevertheless, in practice, predicted team composition (both in terms of demographics and prior arrest and other behavior) also appears balanced across days when Black or Hispanic LTs are predicted to be on duty or not.[6] As a result, any observed differences in team outcomes across days when a Black or Hispanic LT is predicted on duty versus not, are likely not due to differing underlying crime conditions or expected subordinate officer composition.

Using incident-level data on various crime control activities, we construct district-by-watch level outcomes, such as total arrests, for each day. Our reduced-form findings, which are all based off a specification that compares outcomes across days within the same district-watch-year, reveal that on days when Black and Hispanic leaders are predicted to work, subordinate officers make 0.13 (or 3.6% relative to the mean) fewer arrests per day.[7] Arrests of persons of different races decline relatively proportionately, but in practice, since Black people are arrested at far higher rates, the absolute number of arrests of Black people falls most dramatically, accounting for 74% of the total reduction in arrests.

When we explore which types of arrests fall under Black and Hispanic LTs, we see that the vast majority of declines occur for offenses of lower severity. For index arrests, those associated with

[5] Our approach is related to Glover et al. (2017), who compare minority cashier's performance when they are scheduled to work with managers who are biased to when they are scheduled with non-biased managers.

[6] One exception to this appears to be that Black sergeants are more likely to share a day-off group with Black or Hispanic LTs, suggesting some coordination in day-off groups between LTs and sergeants.

[7] Because gender and race are correlated, our baseline specification also controls for any predicted male LT on duty. While this effect is certainly of independent interest, in practice we find no clear differences by gender. We discuss this and differences in other, observable, LT characteristics in Section 5.3. In our baseline specification, we also control for the predicted number of LTs on duty and day of the week and holiday fixed effects.

the most serious offenses, we find a statistically insignificant decline of 0.009 arrests per shift, which is 2% of mean index arrest rates. In contrast, we detect a highly statistically significant, 0.12, or 4% decline, in daily non-index arrests. Among non-index arrests, 88% of the declines come from drug-related offenses (-0.04 or 6% of the mean) and public order offenses (-0.07, or 7% of the mean) — liquor laws, disorderly conduct, miscellaneous non-index offenses, municipal code violations, and traffic violations. In particular, we estimate a 17% reduction in disorderly conduct arrests and a 10% reduction in municipal code violations.[8]

These reduced-form estimates measure differences across whether or not a Black or Hispanic LT is predicted to be on duty. In practice, LTs can and do take leave, could swap days off, and could leave the district-watch. An instrumental variable approach instrumenting for whether or not a Black or Hispanic LT is actually on duty yields a 0.28 decline (or an 8% reduction relative to the mean) in overall arrests per day, but, like Glover et al. (2017), we prefer to focus on the reduced form effect throughout. This is because the exclusion restriction that Black or Hispanic LTs only affect team outcomes on days they actually work is strong. It could be violated, if, for example, officers anticipated that a Black or Hispanic LT will be working and did not realize they were off.

The normative implications of such declines are unclear. A large literature has cautioned against viewing arrests as productivity measures (Lum and Nagin, 2017; Ba and Rivera, 2019; Ba et al., 2021b). In a context where there are significant concerns about over-policing and aggressive policing and associated harms, the observed declines under Black and Hispanic leaders could be welfare-enhancing. Indeed, when we examine Dube et al. (2023)'s pre-registered measure of 'discretionary' arrests — arrests that in their estimation could perhaps have been avoided and handled differently — we find 11% fewer such arrests under Black and Hispanic LTs (p-value 0.055). At the same time, even if many or some such arrests are undesirable, it is unclear whether Black or Hispanic LTs can discourage (or, equivalently in our context, whether white LTs can encourage) one specific subset of arrests without influencing broader aspects of policing, some of which may be desirable. Indeed, we cannot exclude the possibility of significant decreases in index arrests. The 95% confidence interval includes reductions of -0.03 index arrests per day, amounting to 6% of the mean. As such, we cannot take a definitive stance on the desirability or lack thereof of these changes. Our findings, do, however, clearly show that leaders are able to affect outcomes.

When we link arrests to court outcomes, as in, for example, Mas (2006), Rivera (2022), and Weisburst (2022), we see no clear change in the guilty to non-guilty ratio or take-up to non-take-up ratio of arrests, both overall or within sub-categories. This suggests that although Black and Hispanic

[8]We find no clear effects on use of force, or complaints, but results are noisy. We see no clear change in stops, which could be viewed as one measure of officer activity, which decline by a statistically insignificant 1.4%.

LTs may be able to generate reductions in certain categories of arrests that generally have lower rates of eventual conviction, they do not appear able to generate targeted quality improvements within these categories.

We subject these headline results to extensive robustness checks, including a bevy of additional controls, a randomization inference approach and associated placebo check, estimating district-by-watch-specific effects, poisson pseudo-maximum-likelihood estimation, and alternative ways of specifying the treatment. Results are insensitive to these alternatives. We also estimate how effects vary by other observable LT characteristics, such as gender and years of experience. Including these additional dimensions does little to affect our baseline estimates, and, with the possible exception of tenure, reveals little additional heterogeneity across LT types. This suggests that the bundle of experiences and characteristics linked to an officer race are not well proxied for by other observables.

What drives the observed declines in arrests? They do not appear to be due to differential underlying crime conditions or predicted and actual changes in team composition.[9] These factors are generally balanced across the relevant sets of days, and controlling for them changes very little. Nor are effects the likely consequences of differences in policing strategy or patrol behavior, as we see no difference in police stops overall or across neighborhoods, and this would in any case fall outside the typical role of the LTs we study.

Instead, we find that arrest declines appear strongest for Black and Hispanic LTs we identify as watch operations LTs. Effects for field LTs (or, in a separate identification approach) and tactical LTs, appear closer to zero. This suggests that effects are driven by implicit or explicit direction from the district headquarters, as opposed to leaders' actions in the field. While we cannot definitively pin down exactly which watch operation LT practices/attitudes/actions (or officers' perceptions thereof) drive the marginal reductions in arrests, several points reinforce the idea that differences in watch operation LTs practices, broadly speaking, plausibly drive our results. First, watch operations LTs roles make them responsible for approving arrests. They can exert pressure on officers to meet certain standards, make it difficult to obtain approval for certain arrests, and potentially, in extreme cases, even refuse to approve an arrest. Second, we also find evidence indicating that Black and Hispanic watch operations LTs attribute less significance to lower-level arrests than their white counterparts when proposing awards for arrests, suggesting divergent preferences in the approach to arrests. Lastly, we show that our estimated effects strengthen over the course of the year, as subordinates have more contact with supervisors, consistent with what one

[9] As an aside, using variation in the share of officers who are Black and Hispanic purely arising from day-off group variation, we can roughly replicate Ba et al. (2021b)'s main finding that minority officers make fewer arrests. We also document a separate, independent arrest-reducing effect (albeit of sizably smaller magnitude than our LT effect) for minority sergeants.

would expect if it takes some time for an LTs policies/practices to become known.

Understanding the role of supervisors is historically a hard problem due to data limitations and identification challenges (Bloom and Van Reenen, 2011). Outside of the police setting, there has been a recent flourishing of empirical research tackling these challenges (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, 2023). We join these papers in emphasizing the potential importance of managers in the police setting, where incentives and oversight may differ from private firms and where managers influence activities with direct social implications.

Most directly, we relate to a large literature on the effects of diversity in police settings (Donohue III, John and Levitt, 2001; Anwar and Fang, 2006; McCrary, 2007; Ba et al., 2021b; Goncalves and Mello, 2021; Miller and Segal, 2019b; Hoekstra and Sloan, 2022; Rivera, 2022). While these studies generally focus on how minority *officers* police differently, this paper leverages the quasi-random rotation of *supervisors* to explore the influence of supervisors of different races. Our findings, which show that police leaders can impact arrest levels, speak to the importance of leadership and supervision. As such, we complement other, recent work documenting that supervisors can play important roles. For instance, Rim et al. (2022) find that white police supervisors are less inclined to nominate Black officers conditional on work performance than white officers, and Frake and Harmon (2023) find that officers working under supervisors with a high complaint history are more likely to receive complaints. Differently from these papers, our research directly links supervisors' racial diversity to changes in day-to-day shift outcomes. Future work and more data could help move beyond using race as a proxy for a different bundle of characteristics and help identify which LT characteristics, attitudes, and practices matter most in shaping team-level arrests.

Two recent papers have focused on the role of training in affecting use of force and/or discretionary arrests (Adger et al., 2022; Dube et al., 2023). These interventions target all officers, and hence can be quite costly. Our work, which emphasizes the influence police LTs have, suggests that interventions targeted at supervisors could provide a valuable and potentially cost-effective complement to these broader approaches.

The rest of the paper proceeds as follows. In Section 2, we discuss the CPD, the role of LTs, and the day-off group calendar. Section 3 outlines our data and describes our sample. Section 4 outlines our empirical strategy, and Section 5 presents results, robustness, and discusses key takeaways. Section 6 concludes.

2 Institutional Background

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, shift- (i.e. district-by-watch) level (e.g. total arrests on a given day within that district-watch).

The 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 (71% when excluding administrative functions).[12] Sergeants serve as immediate supervisors who work closely with police officers to ensure the adherence to all procedures and policies. They supervise, train, and coordinate the activities of officers, particularly at crime scenes. Additionally, sergeants serve as liaisons between upper management and officers in the field. Within district personnel, approximately 9% hold the rank of sergeant.

Each district is headed by a district commander, who is "accountable for the performance of all district personnel and the actions taken to address police service and crime suppression activities within a district" and an executive officer, a captain who, designated as second in command, "assists the district commander in areas of operation, administration, strategic planning, and eval-

[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, 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.

uation (Chicago Police Department Special Order S03-03: District operations)." While studying the role of police commanders and captains would be valuable, it is empirically challenging due to the unique nature of their positions within the police structure. In each district, there is only one commander and one captain who typically work Monday-Friday (commanders) or Tuesday-Saturday (captains), diverging from the regular day-off group schedules followed by other ranks. This arrangement poses a challenge, as there is no straightforward within-district control group for these roles: their days off typically fall on weekends which can differ from weekdays for other reasons.

Instead, this paper focuses on the next rank below commander and captain but above sergeants — that of police lieutenant (henceforth LT). Under the commander's guidance, 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, all of whom have wide discretion in managing personnel and operations. The role of watch operations LT must be staffed at all times as the watch operations LT oversees personnel/staffing, the review and approval of all arrest reports and complaints, and others duties as outlined in the CPD directives (see Figure 1). The watch operations LT oversees operations from the district's police station. This role effectively comprises 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.

Additionally, each district employs field LTs who are tasked with responding to and assuming command of major field incidents within their respective districts. It is worth noting that these field LTs can assume the role of watch operations LT if the position is vacant (such as on a day off). Similarly, LTs who typically serve as watch operations LTs can also fulfill field LT roles if needed. While watch operations and field LTs are responsible for all officers within the district during their working hours, tactical LTs specifically oversee the officers in tactical teams. Figure A.2 demonstrates that a significant number of LTs divide their workdays between serving as watch operations LTs and field LTs, whereas LTs assigned as tactical LTs do not commonly assume the role of watch operations LTs.

In this paper, we primarily focus on watch operations and field LTs and do not extensively consider tactical LTs. The reason for this omission is that tactical LTs do not adhere to the standard three-watch schedule; instead, they operate on a separate and less predictable schedule. As a result, it is challenging to incorporate their predicted schedules into our baseline empirical strategy.[13]

[13]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. We develop a secondary strategy that allows us to look at effects across all three LT roles including tactical LTs at the daily district level (as opposed to daily district-by-watch level).

Given the substantial overlap in the duties performed by watch operations and field LTs, we treat these roles interchangeably in our baseline specification. Specifically, we will explore how district-watch level outcomes such as total arrests vary with an indicator for whether any Black or Hispanic watch operations or field LT is expected to be working on the given day (where predicted work schedules are based on the CPD operations calendar as we outline below). Field LTs working in the same district-watch can take over as watch operations LTs on watch operations LTs' days off, so it is important to account for them and not just the watch operations LT. We also consider specifications that separate out these two roles.

Altogether, there are several advantages to studying the role of LTs in affecting district-watch level outcomes: i) they are the highest-ranked position that has to be filled on every watch, ii) their role overseeing personnel and arrests is important (particularly so for the watch operations LT), and iii) there are more LTs than officers at higher ranks which creates variation in *which* LT is working during a given watch that lends itself to empirical study.

2.2 CPD Operations Calendar and Day-off Groups

Our identification strategy exploits variation in predicted LT work schedules (determined at the start of the year) arising from features of the CPD operations calendar. We discuss these features here.

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 6 primary 8.5 hour day-off groups (these groups are named 61, 62, ..., 66).[14] Officers in these groups adhere to a set 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.

For concreteness, Figure A.3 presents the operations calendar for the year 2015. 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 the 6th and 7th in the first full week of January. In the subsequent week, they would be be off on the 12th and 13th, and so on, cycling through this pattern until December, concluding with days off on the 26th and 27th. LTs in any other day-off group will have different scheduled days off. 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-

[14] 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.

adjacent day-off groups (e.g. 62 and 64) will work 2 out of every 6 days together. So while each officer will work with all other officers (and all LTs), they will work with some more frequently than others. This calendar creates pre-determined, rotating variation in work schedules for a full calendar year.

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. The district commanding officers (the watch operation LT and his or her superiors) distribute officers across day-off groups, respecting officers' seniority and preferences, but most importantly, making sure that balance in scheduled manpower is maintained across days. This means that officers, and especially the LTs that are the focus of this study, could influence their day-off group. For example, LTs could theoretically choose their day-off group to 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 Years. 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. Finally, even LTs are subject to staffing considerations – for example, if there are only two watch operations/field LTs in a particular watch, they will not be given the same or adjacent day-off groups.

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.

Over the course of the year, it is possible for officers to change their day-off groups. 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, elective time off) causing potential imbalances in day-off groups. 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. Despite these changes over the year, the day-off groups that are assigned in the beginning of the year tend to be relatively stable. We support this claim empirically when discussing the first stage in Section 4.4.

3 Data and Sample

Data. We use personnel data from CPD, obtained by submitting numerous Freedom of Information Act (FOIA) requests. We focus on the period from 2013-2020, a timeline chosen due to significant organizational changes in the CPD around 2012. 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 Lieutenant (LT) was known as watch commander, a position that could be occupied by both captains and lieutenants. Post-restructuring, the CPD differentiated the roles of captains and lieutenants, assigning captains as executive officers and lieutenants with the direct management of district operations. We end in 2020 as most of our data ends then.

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 actual daily attendance and detailed assignment data that allows us to observe unit, hours worked, the reason for any absence (e.g. scheduled day off or personal leave) and assigned beat and car for each police officer.

Additionally, we also obtained historical day-off group calendars, which play an important role in our empirical strategy. For each day from 1972 to 2021, they indicate which day-off groups are taking their days off from work. We use these calendars in January of each year and assignment patterns in that month to infer each officer's scheduled days of work for the rest of the calendar year.

Our key 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 categorize each arrest as index and non-index. We also examine arrests in each FBI code separately. 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 such as homicides, sexual assault, robbery, aggravated assault and battery, burglary, larceny, motor vehicle theft, and arson. Non-index arrests include all the other arrests that are less severe. We also generate 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. Last, we also use Dube et al. (2023)'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., 2023).

Our court docket data come 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. 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 2013-2017 (instead of 2013-2020) because we need follow-up time to observe conviction.

We also have several additional outcomes, all spanning the entire duration of our study period (2013-2020). These are: use of force incidents, complaints, stops, and 911 calls. Use of force incidents come from tactical response reports. CPD officers are required to fill out tactical response reports whenever force is used. The data contains officer demographics, officer actions, subjection actions, incident time, and location. For each complaint, we observe the date of the incident, the type of allegation, and the accused officer. The two most common types of allegations were excess policing and personnel violations. For each stop, we observe the date of the stop, the officer responsible for the stop, descriptors of the individuals who were stopped, as well as information on the time and location of the incident. For each 911 call registered with the CPD, we observe brief descriptions of the event, incident location, and time.[15]

Our main 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 time stamps). Since we reserve January for assigning each officer their day-off group (as explained in the next section), 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 one day of district-by-watch level outcomes. The data encompasses 63 district-watch combinations covering 504 district-watch-years.[16] On average, 1.3 lieutenants are on duty during any given period, and 20% of these are either Black or Hispanic. The typical district-watch has approximately 3.5 arrests, with an average of around 0.5 index arrests and 3 non-index arrests per day.

[15] In Section 5.3, when we assess how LTs reward various types of arrest performances, we utilize an additional awards dataset. This contains information for each award, including the requestor, the awardee, and the date of the incident for which the award was given. We link it to officer-day arrest records spanning 2013 to 2020.

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

In our sample, 348 lieutenants work across the various district-watch combinations as watch operations or field LTs. Out of these, 77 are Black or Hispanic leaders. Comparing Black and Hispanic leaders to their white counterparts on various metrics such as gender, tenure, awards, and complaints, in Table A.1, we find no significant disparities.[17] However, Black and Hispanic lieutenants are substantially more likely to have been promoted more recently (after 2004) than their white counterparts.

4 Empirical Strategy

4.1 Ideal Experiment

We are interested in how district-by-watch level outcomes such as total arrests differ by self-reported LT race. The randomized experiment we seek to mimic would take, for example, a district-watch with a Black and a white LT and randomly assign each LT to work different days of the year, allowing us to compare daily district-watch outcomes across LT race within the same calendar year.[18] Random assignment is important since it ensures comparability of expected underlying crime conditions and subordinate officer characteristics across days. In practice LTs can take leave, move to a different district, watch, or day-off group, or get promoted, and these actions could be associated with both team performance and potentially race. For example, if Black or Hispanic LTs (relative to white LTs) are less likely to take leave on high-crime days, are called in on more challenging days, or are reassigned to tougher district-watches mid-year, one would observe higher team level arrests under minority LTs.

In this idealized experiment we would primarily be interested in the intention-to-treat effect — that is, being assigned certain days regardless of actually working all of those days. This is because the decision to actually show up to work and not take leave is potentially endogenous and because some decisions that watch operation LTs make, for example, which sergeant they designate to serve as the district station supervisor, could affect outcomes even if they take a day off.

In this paper we approximate this idealized experiment by leveraging features of CPD's operations calendar. Specifically, as outlined in Section 2, at the start of each calendar year, LTs (like other patrol officers) are assigned to one of six day-off groups. This creates a pre-determined set of days that these LTs are supposed to work throughout the year. They can of course still take leave or

[17] In this table, we exclude 8 Asian/Pacific Islander (API) LTs. In our main analysis, they take on the value of 0 in our main regressor for Black and Hispanic leaders. Our results are robust to excluding these few API leaders.

[18] As with all such studies, comparisons across people racialized as Black or Hispanic versus those not could reflect a myriad of differences in inter-group characteristics, behaviors, and life experiences, as well as socio-contextual factors that influence one's identity in the first place.

potentially alter their schedules in other ways, but by only leveraging variation in predicted days at work coming from the operations calendar, we are able to abstract from this potentially more endogenous variation. Helpfully, since work schedules do not overlap with normal weekdays, the set of days a given LT is scheduled to work throughout the year are unlikely to have different average crime rates than the set of days a different LT in a different day-off group is scheduled to work — a claim we support empirically below.

4.2 Constructing Predicted Work Schedules

We are interested in how daily district-watch level outcomes, such as total arrests, vary with the race of the LT(s) predicted on duty within a calendar year. Once we know a LT's district-watch and their day-off group, we can use the operations calendar to construct this LT's predicted work schedule for the entire year (i.e. all the days in that year except for the days off). Repeating this for all field and watch operation LTs initially assigned to that district-watch allows us to construct an indicator for whether any Black or Hispanic LT is predicted to be working in the district-watch on each day — *Any Predicted Black/Hispanic LT*. This will be our key independent variable of interest. In a district-watch with one Black or Hispanic LT and one white LT, this dummy variable equals 0 on the Black or Hispanic LT's days off (when the white LT will be working alone) and equals 1 when either the Black or Hispanic LT is working alone or when the Black or Hispanic LT and the white LT are both working. In district-watches with only 1 identified LT who is also Black or Hispanic, this will equal 0 on their day offs, when a station sergeant (or a substitute LT from another district) will fill in for them.

Two complications in our data are i) we do not directly observe the district-watch to which officers are assigned and ii) we do not directly observe officers' day-off groups. Instead, we need to infer both of these from attendance and assignment records. Our data allow us to see the watch and unit the officer is assigned to on any given day. Additionally, reason-for-absence codes provide a specific code for a scheduled 'day off'. Hence, we take all records from January, assign each officer to their modal district and modal watch and to the day-off group that best matches their monthly schedule. As we show below, this imputation is very accurate. We correctly identify over 97% of actual scheduled days off (i.e. absence code is 'DAY OFF') in January (both for LTs and officers). Identifying an officer's district-watch using their modal district watch also works well, with officers actually working in their modal district-watch on over 98% of days in January. Given that we use January to impute day-off groups and assigned district-watch, our results will focus on daily, shift-level outcomes between February-December.

4.3 Baseline Specification: The Reduced Form Effect of Any Predicted Black or Hispanic LT

Our empirical approach compares 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. All daily comparisons are made *within* a district by watch by year. Formally, we estimate the following reduced form specification for district-by-watch-by-year j on day t.

$$\begin{aligned} y_{jt} = \beta_1 &\text{Any Predicted Black/Hisp. LT}_{jt} \\ &+ \beta_2 \text{Any Predicted Male LT}_{jt} + \beta_3 \text{Predicted Num of LTs}_{jt} + \alpha_{f(j)} + \delta_{g(t)} + \epsilon_{jt}. \end{aligned} \tag{1}$$

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. 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 is 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). Unless we note otherwise, we also use the following baseline fixed effects in all our regressions: $\alpha_{f(j)}$ includes district×watch×year fixed effects as well as separate day-of-the-week fixed effects and separate fixed effects for all city of Chicago holidays ($\delta_{g(t)}$). These help absorb common shocks to underlying crime rates and help alleviate concerns around LTs choosing day-off groups around any one particular holiday. We cluster standard errors at the district-watch-year level. 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.

Our focus throughout the paper 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 find no statistically significant differences across gender (as shown and discussed in Section 5.3). In that section, we also probe how outcomes vary with other, observable LT characteristics besides gender and race.

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. Restricting to teams with variation in *Any Predicted Black/Hisp. LT*

makes little difference for both the coefficients and standard errors as we show later.

This reduced form specification shows how team level daily, shift-level outcomes like arrests vary across days when a Black or Hispanic LT is predicted to be on-duty. In cases when the watch operations LT is off duty, this position is typically filled by the field LT or, if no field LT is available, the district station sergeant (or a substitute LT from another district). In cases where a field LT is off duty, the position need not necessarily be filled, but the watch operations LT may respond to a field incident. In practice, in the majority of cases, the counterfactual involves having one or more white LTs actually on duty.[19]

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 only allows us to identify relative effects. For example, if we see a reduction in arrests, we cannot tell if this is because Black or Hispanic LTs discourage certain types of policing or if, say, white LTs incentivize or reward these types of policing.

4.4 Balance and First Stage

Balance on expected crime. Since our identifying variation ultimately arises from the day-off group an LT is assigned to, and LTs could potentially influence their day-off group, 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. This assumption appears reasonable because days off for each group rotate throughout the year, and, after accounting for major holidays, it is challenging for LTs to accurately predict differences in crime patterns far into the future. Nonetheless, we probe these claims empirically.

Using the specification in Equation 1, we explore in Table 2 whether days with any predicted Black or Hispanic LT look systematically different from other days in terms of expected crime.[20] Panel A shows that daily arrests in other police districts are uncorrelated with whether a Black or Hispanic LT is expected to be on duty in the given district. This remains true if we focus only on arrests in nearby districts (same CPD area), arrests in all other districts during the same watch,

[19] 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 5.2, we also present estimates using a binary indicator for any predicted white LT and find roughly equal and opposite results.

[20] We transform the dependent variables using the inverse hyperbolic sine, but Table A.3 shows that our conclusions are unchanged when leaving the dependent variables as raw counts.

and arrests in only nearby districts in the same watch. Panel B shows that 911 calls in the given district-watch are not correlated with the race of the LT who is expected to be working. Since 911 calls could at least theoretically be influenced by who is in charge, subsequent columns of panel B also show that daily 911 calls in all other districts are balanced across race of the given district's LT. In panel C, we predict arrests using all the variables from panel A (and our baseline controls), and then use predicted arrests as our outcome of interest in Equation 1. Predicted daily arrests are also uncorrelated with our 'treatment' variable. In subsequent columns, we first add squares and cubes of leave out arrests (column 2), then all the 911 call variables from panel B and their squares and cubes (column 3). Column 4 additionally uses patrol officer and sergeant demographics (as expected based on their day-off groups). Regardless of specification, predicted arrests are uncorrelated with whether a Black or Hispanic LT is expected to be on duty. Overall, available evidence suggests that Black and Hispanic LTs are just as likely as white LTs to have days off on high crime days.

Team composition. Another aspect that could differ in the work environment of teams managed by Black and Hispanic LTs is the composition of subordinate officers. The composition of these teams may vary for two primary reasons. First, certain types of officers may be more likely to be in day-off groups that have more overlap with Black and Hispanic LTs (as might be the case if officers tried to coordinate on their first choice day-off group). 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, while initial day-off group assignments of officers aim to maintain a balanced distribution among teams, deviations in actual work days can occur due to factors such as transfers. Third, selective work attendance through use of leave or injuries could alter actual group composition. Hence, it is useful to briefly explore how both predicted teams (based on initial day-off groups) and actual, eventual teams, differ across days with a predicted Black or Hispanic LT.

In Table 3, we explore the correlation between having a predicted on-duty Black or Hispanic LT and both predicted and actual officer composition. First, in panels A-C we look at the correlation between having a predicted Black or Hispanic LT and predicted officer 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' arrest and use of force records in the prior year also appear statistically comparable across days when Black or Hispanic LTs are predicted to be working. Panels B-C separate results out by sergeant versus 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 (as shown 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 sergeant characteristics. Out of the 24 team characteristics we study, this is the only one that is statistically significant. Since one of the roles of the watch operations LT is to designate a sergeant to serve as the district station supervisor and since they are also effectively in charge of day-off groups, this suggests that LTs might be handpicking their sergeants.[21]

In panels D-F, we explore the correlation between having a predicted Black or Hispanic LT and *actual* officer and sergeant demographics during that day and watch. While coefficients tend to not differ statistically from their counterparts in panels A-C, they move in the direction one would expect if watch operations LTs enabled some degree of same-race sorting when deviations from initial day-off groups occur. We see more Hispanic officers and more officers with lower arrest histories on days when Black or Hispanic LTs are predicted to be working. However, since overall magnitudes remain small and changes from predicted demographics are not significant, it appears that any active management changes to teams (e.g. moving day-off groups) is at most limited in how it affects the racial mix of teams.

Altogether, this table suggests that some Black or Hispanic LTs might hand-pick a sergeant to work on the same four days as themselves, but there is little evidence that Black and Hispanic officers are more likely to be pre-assigned to the same day-off groups as Black or Hispanic LTs. Actual, as opposed to predicted, team composition shows hints of same-race matches between Hispanic officers and minority LTs, but ultimately any differences across treatment and control groups remain quantitatively small. As a result, in Section 5, we will show that controlling for these demographics (predicted or actual) makes little difference for our estimates. The effects of having a Black or Hispanic LT on team-level outcomes are not likely operating through the racial or other observable makeup of the team.

First stage. Finally, we turn our attention to the effective first stage — specifically, evaluating how often between February and December, on days we predict a Black or Hispanic LT to be on-duty, they are indeed observed on duty. Prior to estimating the first stage, it is useful to assess the stability of day-off groups and district-watch assignments, and to understand how leave patterns affect our predictions.

In January of all years we correctly predict 98% of days off for both officers and LTs. From

[21] 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. However, we later show that controlling for predicted or actual sergeant composition does little to results. This suggests that our results predominantly capture the impact of LTs. Nevertheless, teaming up with certain sergeants over others could be one mechanism through which LTs exert influence.

February-December, we correctly predict 88% of days off for officers and 85% of days of for LTs. This match rate declines over time but remains over 79% (see Figure A.4). Thus, day-off groups are relatively stable throughout the year. However, LTs could also move watches or districts (or be promoted or leave the force). In January, 99% of days for both officers and LTs are spent in the district-watch we identified them in. From February-December, 91% of officers and 81% of LTs are in the district-watch we initially identified them in. As a result, these two forces (day-off group and district-watch stability) imply that on 84% of days for officers and 77% of days for LTs between February and December we both see the officer in their initial district-watch and correctly predict their days off.

Predicting scheduled days off and predicting days actually present is not the same thing. All officers and LTs take vacation on top of their days off, for example. More generally, absences on any non-day off could be due to vacation, illness, leave, or day off compensatory time, among other reasons. As a result, when we examine the percentage of days we correctly predict a LT is at work based solely on day-off groups *and* only count it as getting it right if they are in the same district-watch as we identified, then we get it right on 65% of the days (72% for officers). Altogether, prediction errors arise because of day off group changes, no longer being in the original district-watch, not working on a scheduled day (due to things like vacation), or potentially working on a day off.

With this context, we present our first stage estimates of the relationship between our prediction of when Black or Hispanic LTs will work and when they actually work. Among the set of teams that provide our identifying variation (i.e. district-watch-years for which our binary variable turns on and off over the course of the year; in practice, these are district-watches in which we identify a single Black or Hispanic LT working in January), we correctly predict whether a Black or Hispanic LT is working on 68% of days. This is in line with the 65% described above. These predictions correspond directly to a first-stage coefficient (i.e. using "Any Black of Hispanic watch operations or field LT on duty" as the outcome in Equation 1) of 0.46 (Table A.2).[22]

This first stage coefficient is far higher than what you would get by chance. As a point of comparison, consider the district-watch-years that provide our identifying variation, and imagine assigning

[22]To understand this better, we can cross-tabulate our binary prediction of Black and Hispanic LT presence with reality. We calculate this only within district-watch-years where a Black or Hispanic LT was operating in January because these provide the key identifying variation in our model. Other district-watch-years with no variation are absorbed by fixed effects. On 29% of all days, we say there is no minority LT, and we are correct. On 4% of days, we say there is no minority LT, and we are wrong. On 39% of days, we say there is a minority LT at work, and there is one. Finally, on 28% of days, we say there is a minority LT at work, but they are not (e.g. vacation). Overall, we are correct on 68% of days (29+39). In the simple case where there are no additional controls besides the district-watch-year fixed effect, this cross-tabulation implies a first stage coefficient of 0.459. Since, in the single binary independent variable case, $\hat{\beta} = E(Y|x=1) - E(Y|x=0) = \frac{39}{39+28} - \frac{4}{29+4} = 0.459$. The additional controls for LT gender, number of LTs, holidays and day of week FE change the estimate relatively little in practice.

each Black or Hispanic LTs in these district-watch-years to a randomly selected day off group (or, analogously, to work on any random 4/6 of days). On average, this yields a first stage of zero. Indeed, we later leverage a randomization inference (and placebo) check in which we do precisely this. The average first stage across the 2000 trials is -0.0007 (se = 0.0275).

For the rest of the paper, we could of course report an IV specification for actually having a Black or Hispanic LT on duty (and we do in Table A.4), but we prefer to focus on the reduced form since it is possible that a minority LT could influence outcomes on days in which they are supposed to be on duty but take leave (for example if officers expect them to be on duty and don't realize), violating the exclusion restriction. A similar approach of focusing on reduced form results has been used by Glover et al. (2017).

5 Results

5.1 Fewer Arrests under Black and Hispanic Leaders

Do police teams perform differently under Black and Hispanic leaders? Table 4 presents regression estimates of Equation 1 for various arrest measures, comparing daily district-watch level outcomes on teams with a predicted Black or Hispanic watch operations or field LT to those without. Notably, having a scheduled Black or Hispanic leader results in a statistically significant drop in overall arrests by 0.13, which accounts for a 3.6% reduction relative to the mean of 3.5 arrests. Arrests fall relatively proportionately across race of the person being arrested, with a 3.8% decline in arrests of Black persons (column (2)), a 2.8% decline in the arrests of white persons (column (3)), and a 3.0% decline in the arrests of Hispanic persons (column (4)), all expressed relative to the corresponding mean. However, since arrests of Black persons are far more frequent, the absolute number of arrests falls most dramatically for this group, accounting for 74% of the total reduction in arrests.

Instrumental variable results are presented in Table A.4. Scaling the reduced form by the inverse of the 0.46 first stage in Table A.2, yields a decrease of 0.28 arrests (or an 8% reduction relative to the mean). These results require stronger assumptions to interpret causally. In particular, IV requires an additional monotonicity and exclusion assumption, the latter of which requires that having a predicted Black or Hispanic LT only affects team outcomes on days they actually work. This could be violated, for example, if officers anticipated them to be working but didn't know they were off or if officers who usually work with them have adopted certain attitudes that persist under a different one-off, substitute LT. Hence, we prefer to focus on the more conservative reduced form effect.

Figure 2 and Panel B of Table 4 explore which types of arrests decline under the leadership of Black or Hispanic LTs. Figure 2 presents results for arrests in each FBI code, separately, as well as for several groupings of arrest categories. We see no statistically significant effect on index arrests, which decline by 2% of 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 0.119 or 4% of the mean. When examining each separate FBI code, it becomes clear that non-index arrest reductions come primarily from what we term public order offenses (colored in green) — liquor laws, disorderly conduct, miscellaneous non-index offenses and municipal code violations, and traffic violations — as well as, perhaps, from drug-related arrests. We estimate a 0.03 (17%, p-val 0.002) reduction in disorderly conduct arrests, a 0.01 (4% , p-val 0.15) reduction in miscellaneous non-index offenses, a 0.01 (10%, p-val 0.08) reduction in municipal code violations, a 0.02 (5%, p-val 0.05) reduction in traffic violation arrests, and no change in liquor law arrests. As a result, when we group these categories together in Table 4, we see a strongly statistically significant decline in public order arrests by 0.07 (7%). Drug arrests decline by 0.04 (6%, p-val 0.052). Together, public order arrests and drug arrests account for over 88% (0.112/0.127) of the total, observed reduction in arrests.

The normative implications of these arrests declines are unclear and depend on whether or not the arrests reduced at the margin are productive or societally undesirable. As in much prior literature, we caution against viewing declines in arrest as a measure of productivity (Lum and Nagin, 2017; Ba and Rivera, 2019; Ba et al., 2021b). Indeed, based on the type of arrests that we see declining — largely minor violations and drug-related offenses, but not index arrests — it is possible that the types of arrests being reduced are closest to what Dube et al. (2023) 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 plausibly unnecessary)". Their pre-registered arrest grouping for discretionary arrests — intended to capture arrests that most likely fit this description but not necessarily all such arrests — contains many of the arrests for disorderly conduct that we see decline the most. In fact, when we use discretionary arrests, constructed using their definition, as an outcome in column (4) of Table 4, we find that having a predicted Black or Hispanic LT reduces discretionary arrests by 0.0074 arrests (standard error 0.0039; p-value 0.055) off a mean of 0.068 (or 11%). The training program studied in Dube et al. (2023) reduces these arrests by 23%, so the reduced form effect of having a predicted Black or Hispanic LT on duty appears substantial.

While this suggests that some of the observed declines under Black and Hispanic leaders could be indicative of welfare-enhancing change, we urge caution. It is not clear that Black or Hispanic LTs are able to discourage (or, alternatively, that white LTs are able to encourage) one specific subset

of arrests without affecting broader, potentially helpful, arrests. Indeed, we are unable to rule out potentially meaningful declines in index arrests. A 95% confidence interval rules out declines larger than 0.03 arrests per day, which corresponds to 6% of the mean.

We expand on this discussion in Sections 5.3 and 5.4, where we link arrests to court outcomes and explore other outcomes, among other exercises. Before doing so, however, we conduct several robustness checks to reinforce our main findings.

5.2 Robustness and Specification Checks

We discuss several robustness checks in turn. First, we include additional controls. In Table A.5, we first add fixed effects for day of the year (e.g. Jan 1) in the event they better capture underlying crime trends. In Panel B, we control for the leave out arrests and 911 call variables used in panels A and B of Table 2. In Panel C, we control for predicted team composition (Panels A-C of Table 3). These additional controls all do very little to our baseline results (as does including all these controls simultaneously, as shown in Panel D).

One additional, noteworthy, result, is that when controlling for the predicted daily share of subordinate officers who are Black or Hispanic, based on day-off group variation, we see fewer arrests when there are more minority officers (Table A.6). Specifically, a 1 standard deviation increase in the share of Black officers (=0.18), or approximately 10 more Black officers, is associated with a 0.1 decrease in daily arrests arrests ($0.18 \cdot -0.541$). Over 100 shifts, this means 10 more Black officers are associated with 10 fewer arrests, or about 1 arrest per officer. This is in line with Ba et al. (2021a), who find that Black officers make 1.93 fewer arrests than white officers per 100 shifts. The analogous calculations for Hispanic officers in our sample yield about 0.8 fewer arrests per 100 shifts per Hispanic officer (as compared to around 0.5 in Ba et al. (2021a)). As such, we are able to broadly reproduce their main findings in our framework, exploiting only variation induced by day-off group assignments.

Table A.7 shows that restricting to the 126 district-watch-years that exhibit variation in having a predicted Black or Hispanic LT (i.e. excluding district-watch-years with no predicted Black or Hispanic LT ever and district-watch-years that always have a Black or Hispanic LT predicted to work on each day), yields statistically indistinguishable results from our baseline. If anything, point estimates are now slightly larger in absolute value.

Among these same 126 district-watch-years, Table A.8 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 counter-factual for when a Black or Hispanic LT has a day off in this sample is in-

deed having a white LT predicted on duty. Note that the estimated coefficients are not exactly equal and opposite because of overlapping days worked and because in some teams the counterfactual could be having no predicted LT on duty or an Asian/Pacific Islander LT.

In Table A.9, we consider 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. This is done 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 randomization inference procedure yields standard errors close to those obtained when clustering at the district-watch-year level. 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. Standard errors when clustering at the district-watch level (which account for potentially correlated errors in the same district-watch over time) are generally larger, as one would expect, but not sufficiently different as to alter our main takeaways.

In Figure 3, we consider whether results are driven by specific district-watch-years (and, hence, in turn, by a few select LTs) or whether results hold more broadly. 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 1). To account for noise, we shrink the estimated treatment effects using empirical bayes, but this does not substantively alter the main takeaway. The distribution of district-watch-year specific treatment effects is centered around our main treatment effect estimate of -0.127 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 -0.127, one quarter of district-watch-years have a positive treatment effect larger than 0.127. 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.

Because arrests are count variables, we also show that results are highly similar when using a poisson pseudo-maximum-likelihood estimator in Table A.10.

While we initially considered both Black and Hispanic lieutenants together, Table A.11 Panel A presents a separate examination of the results based on the lieutenant's race. The point estimates show that team arrests decline under both Black and Hispanic lieutenants, although the measure-

ments are imprecise. To help with precision, Panel B adds controls for leave out arrests, 911 calls, and predicted team composition. Overall, declines in arrests look similar across Black and Hispanic LTs.

Last, instead of our binary specification for any Black or Hispanic LT predicted on duty, in Table A.12 we use the share of LTs who are Black or Hispanic. In practice, this is effectively 0, 0.5, or 1 (it is $1/3$ and $2/3$ for 0.43% of observations), with 0.5 and 1 about equally likely when a Black or Hispanic LT is predicted on duty. This exercise yields similar conclusions.

5.3 Contextualizing Arrest Declines and Exploring Potential Channels

In order to better contextualize our results and explore potential channels, we present a series of additional results. We first estimate effects on other outcomes (e.g. use of force), link arrests to court outcomes, and explore the effects of other, non-race based, LT characteristics on outcomes. We then probe potential channels, including differential team composition or patrol patterns, and which types of supervisory activities (e.g. in field vs at headquarters) appear to matter most. We show that watch operations LTs, who oversee arrests, drive our effects and discuss evidence consistent with the idea that their supervisory standards and practices regarding arrests are plausible drivers.

Other outcomes. Table A.14 explores other outcomes: total stops, total use of force incidents, and total complaints. We do not detect any statistically significant effects of Black and Hispanic LTs on these outcomes, though we are unable to rule out quantitatively meaningful changes in use of force and complaints. For example, the 95% confidence interval for use of force incidents ranges from an 11% decrease relative to the mean to a 5% increase relative to the mean, with the point estimate suggesting a 3% decrease. We are similarly under-powered to detect changes in complaints and complaints relating to use of force. Stops, which could perhaps be viewed as a measure of activity, decline by a small and statistically insignificant 1.4%.

Court outcomes. One way to examine the potential productivity of arrests is to link them to court outcomes (Mas, 2006; Weisburst, 2022; Rivera, 2022). Because we need follow-up time for conviction outcomes when doing so, we restrict our sample to the years 2013–2017. We construct outcomes along two dimensions: conviction and take-up by the court. In Table A.15, we present results for arrests linked to conviction outcomes. Overall, under Black and Hispanic LTs both arrests that led to conviction and arrests that did not lead to conviction declined. Arrests that led to conviction decreased by 0.083, or 3.8% of the mean, while arrests that did not lead to conviction decreased by 4.7%. To understand how the relative proportion of guilty-to-non-guilty arrests changed, we add the estimated coefficients for guilty and non-guilty arrests to the respective

means and compare the resulting ratio with the ratio of the means (Panel C). For overall arrests, the estimated ratio of guilty-to-non-guilty arrests increases by a statistically insignificant 0.012 relative to the ratio of means (2.207/1.755 = 1.26). Thus, on average, guilty and non-guilty arrests appear to fall relatively proportionately under Black and Hispanic LTs.

When we repeat this exercise for different sub-categories we obtain much the same picture: there is no clear change in the guilty to non-guilty ratio under Black and Hispanic LTs for any sub-category of arrests. Nevertheless, it is worth pointing out that non-index arrests have substantially lower conviction rates than index arrests (about 52% relative to 77%).

Since many arrests do not lead to convictions, it also makes sense to consider whether the court considers the case in the first place. In Table A.16, we examine this alternative margin and find much the same picture. There is no clear change in the take-up to non-takeup ratio for any arrest category. As with the guilty to non-guilty ratio, the take-up to non-takeup ratio is higher for index crimes than non-index crimes.

Altogether, our results suggest that low-level arrests decline under Black and Hispanic LTs, but that these arrests would likely have gone on to the courts and to eventual conviction at roughly average rates had they occurred. This suggests that while LTs may be able to produce declines in certain arrest categories, such as non-index arrests that generally have lower rates of eventual conviction, they do not appear to generate targeted quality improvements *within* these arrest categories, at least based on court outcomes.

Effects are not well captured by other, observable LT characteristics. In Table A.17, we control for other characteristics, beyond gender and race, of the LTs predicted to be on duty. Specifically, we construct binary versions of the variables in Table A.1 (e.g. any Pred LT with more than 20 years of tenure) and include them in our specification. First, we note that these additional controls do not alter the effects of any predicted Black or Hispanic LT, consistent with these variables looking relatively comparable across races in Table A.1. Second, we generally do not see any clear effect of having a male predicted LT on duty. Nor, generally speaking, do we see differential arrests under LTs with more awards, complaints, or LTs promoted more recently. There is some evidence that arrests are lower under LTs with more experience (tenure). Altogether, this indicates that the bundle of characteristics associated with race that drives our results is not well-captured by available, observable histories. It is possible that differences in race emerge in part due to differences in lived experience. In a 2017 investigative report, the Department of Justice notes "CPD's own officers, especially, but not only, its black officers, acknowledge profiling and harassment by CPD. A lieutenant told us, "I'm a black man in Chicago, of course I've had problems with the police"" (US Department of Justice, 2017). This does not imply that results are driven

by immutable characteristics; we expect these lieutenants to have different management practices, leadership qualities, expectations, and other habits that presumably LTs of all races could emulate if desirable – something we discuss further below.

Effects are not driven by differential team composition. Since LTs can play a role in regulating day off schedules and may be picking their sergeants, it is worth considering the extent to which differences in outcomes could be generated by differences in team composition. As shown in Table 3, the predicted composition of officers across days when Black and Hispanic LTs are expected to be on duty versus when they are not appears relatively balanced (with the exception of Black sergeants who are more likely to share a day-off group with minority LTs). Actual officer and sergeant composition varies slightly from predicted and shows hints of increased same-race matches, but differences are small in magnitude. This is consistent with watch operations LTs filling schedule gaps disproportionately with minority officers, and could, since minority officers tend to make fewer arrests, potentially explain a (small) fraction of our results. In practice, though, including (potentially 'bad') controls for all the actual officer and sergeant demographics and arrest histories in Panels D-F of Table 3, does almost nothing to our estimates (see Table A.18).

Interestingly, minority sergeants do exert their own independent effect on arrest reductions. Table A.19 shows that a 1 standard deviation increase in the predicted share of minority sergeants on duty – or approximately one additional minority sergeant — decreases arrests by 0.028 (0.21$\times$ -0.135). As such, the effect is more muted than that of a Black or Hispanic LT but nevertheless meaningful. Controlling for an independent effect of minority sergeants does little to the baseline effect of a minority LT. There is also no interaction effect between these two variables (not shown). Combined with Table A.6, this shows that total district-watch level arrests decline when there are more minority police officers *at all ranks* (officers, sergeants, and LTs). However, sergeants, and, especially, LTs differ from regular officers in that their roles typically do not involve making arrests in the field. Together with the fact that the effects of sergeants and LTs are robust to controls for both predicted and actual district-watch composition of officers, this implies that sergeants and LTs are exerting their influence on subordinate officer behavior through other means — such as setting and enforcing standards for arrests or providing guidance and oversight in the field.

Effects are concentrated on watch operations LTs. Our baseline analysis combines the effects of watch operation and field LTs. To gain a deeper understanding of the roles that management plays, we examine the distinct contributions of these two roles. Field LTs are responsible for on-site supervision, which includes directing responses to major incidents, ensuring patrol coverage, and monitoring and managing patrol activities. On the other hand, watch operations LTs oversee operational and administrative functions, such as reviewing arrests and use of force incidents and ensuring compliance with policy and procedures. Table A.20 explores how effects vary when

considering these two roles separately based on their predominant role in January. Point estimates are generally largest in absolute value for watch operations LTs but we lack statistical precision to formally distinguish between the two roles (p-value = 0.16). Field LTs can act as watch operations LTs if needed, and roles could change over time, muddying the contrast. But these results hint that watch operations LTs, who would spend most of their time in the headquarters as opposed to the field, are playing a key role in our results.[23]

Due to their rotating and flexible hours on any given day, we cannot reliably predict whether a tactical LT is working on any given district-watch. However, an alternative strategy that examines daily district-level outcomes (as opposed to district-watch-level outcomes), enables us to also leverage tactical LT day-off group variation in whether any Black or Hispanic tactical LT is predicted on duty on any given day. The results from this exercise are shown in Table A.22 and reinforce the idea that effects most likely stem from watch operations LTs and less so for tactical or field LTs. The point estimate on having a Black or Hispanic tactical LT is positive and close to 0. The p-value for a test that the coefficient on any Black or Hispanic watch LT is equal to that for tactical LTs is 0.119, and the p-value on a joint test of equality across all three coefficients is 0.081. This suggests that any potential mechanism behind our results stems from influence being exerted at the district headquarters, as opposed to influence while on the scene.

How do watch operations LT influence their subordinates? We propose that it is most likely watch operations LTs are enacting change through their role as the ultimate standard enforcer and reviewer and approver of all arrests, rather than through any role they play in officer assignment to beats or affecting crime strategies. First, watch operation LTs would not typically alter the beat assignment of officers. In the CPD, beat officers consistently patrol the same beat on the same watch for at least a year. This consistent assignment is structured to enable beat officers to cultivate familiarity with the individuals on their beats, thereby fostering relationships between community members and their assigned beat officers. Moreover, assignments of the more flexible tactical officers are overseen by tactical LTs, and not watch operations LTs. Last, the formulation of crime prevention plans and strategies occurs at a level above the LTs, specifically at the executive officer and district commander levels. Table A.23 provides partial empirical support for this by examining stop behavior. It shows that both overall daily stops within the district-watch and stops across different subsets of beats are similar under Black/ Hispanic LTs and white LTs.

Instead, LTs most likely exert influence through their supervisory role in ensuring compliance with arrest policies and procedures. Watch operations LTs can make it challenging for officers who do

[23] An alternative specification using the dummy variable for whether any white operations or field LT is predicted on duty yields near equal and opposite results (Table A.21). Results are also more precise: we can clearly reject the null that the coefficient on white watch operations LTs is equal that for white field LTs, reinforcing the idea that watch operation LTs play a predominant role.

not meet their arrest standards. Anecdotal accounts from our interviews with CPD police officers support the notion that watch operations LTs' standards and attitudes can shape officers' decision-making. 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.[24] For instance, certain watch operations LTs may express disapproval of trivial arrests and imply that these will not be rewarded. They can also make it more challenging or burdensome to obtain approval for such arrests. While rare, LTs could also not approve arrests brought into the station at all (in which case the event would not show up in our data as an arrest). In response to such interactions, officers may adapt their behavior and refrain from making marginal, low-level arrests that may not receive recognition or support.

While lack of relevant data prevents us from definitively pinning down exactly which practices and attitudes (or officers perceptions thereof) differ across LTs of different races, two points reinforce the idea that watch operation LTs' standards and practices regarding arrests plausibly drive our results. First, using arrest and awards data from 2012 to 2020, we examine nominations for awards by both white and Black and Hispanic LTs. LTs play an important role in nominating awards for commendable arrest performance, typically recognizing serious index arrests. However, various factors can impact nomination behavior.[25] If certain LTs also value lower-level arrests, they may be more inclined to grant awards to officers who make a higher number of such arrests, even among those who make the same index arrests. Although awards are infrequently granted, exploring the variability in nomination propensity for different arrest types provides insights into the level of importance LTs attribute to these specific arrest categories.

Using daily arrests data linked with awards data, we investigate how different daily arrest performances are acknowledged by white and Black and Hispanic watch operations LTs. In Table A.24, we find that both index arrests and non-index arrests increase the likelihood of receiving awards from white and Black and Hispanic watch operations LTs. For example, an additional index arrest by an officer on a given day is associated with a 0.00151 increase in white watch operation LT-nominated awards on that same day. The mean number of awards nominated by Black or Hispanic watch operations LTs are about 1/5 of that from white LTs. This difference is similar to the proportion of Black and Hispanic LTs in CPD. However, the relative contribution of non-index arrests to index arrests to the propensity of receiving award nominations differs between white and Black or Hispanic watch operations LTs. For white watch operations LTs, non-index arrests count

[24] Additionally, several public investigations by the Civilian Office of Police Accountability (COPA) show that LTs can face consequences for incorrectly approving arrests or reports (e.g. COPA log 1090147; COPA log 1091065). While rare (the majority of complaints filed with COPA do not sustain evidence against CPD officers), these cases highlight the important role LTs play and hint at how variation in LT standards could lead to disparate arrest outcomes.

[25] For instance, Rim et al. (2022) show using CPD data that officer race can play a role in white LTs' propensity to give awards given the same officer performance.

for 25.2% as much weight as Type 1 arrests, while for Black and Hispanic watch operations LTs, non-index arrests only account for 14.7%. This lower weight given for non-index arrests by Black and Hispanic watch operations LTs in award decisions suggests that watch operations LTs of different races may place differential value or emphasis on low-level arrests. If Black or Hispanic LTs incentivize making non-index arrests less than white LTs, this could help explain why we observe fewer such arrests under Black and Hispanic LTs.

Second, enforcing new standards and expectations takes time. Since the group of officers assigned to work most frequently with any particular LT reset at the start of year due to the day off group calendar, we can ask whether estimated effects strengthen over the course of the year (with more opportunity for interactions between officers and the LT). Recall that day-off groups work in such a way that some officers would overlap with one watch operations LT on all days they work (4 out of every 6 days), while others will do so as few as 2 out of every 6 days. It is plausible, especially given this schedule, that officers need time to get to know their watch operations LT(s). As a result, we might expect stronger effects over the course of the year, as officers get used to the LTs they work with most. Table A.25 shows that when we interact the indicator for any predicted Black or Hispanic LT at work with months since February (measured from 0 to 10), we see that the reduced form effects strengthen over the course of the year. This suggests that the effects of a predicted Black or Hispanic LT get stronger the more time officers have to get used to the LT. The fact that changes take time is also consistent with leaders enacting change (as opposed to immediate knee-jerk responses on the part of officers to an LT's race).

5.4 Discussion

What can and can't we conclude from the preceding set of results? First, arrests clearly and robustly decline under Black and Hispanic LTs relative to days without a predicted Black or Hispanic LT on duty. These differences are not likely due to differing average crime conditions, team composition, or beat assignments.

We propose that the observed reductions in arrests are due to differential standards and expectations of watch operations LTs. Arrest declines stem from the watch operations LT, who is the final approver of all arrests and who is stationed in the district headquarters as opposed to in the field. We find that Black and Hispanic watch operations LTs differentially reward low level arrests relative to their white peers, and that our main effects generally strengthen with contact time. Without more granular data on actual practices, we cannot definitively pin down the exact practices through which LTs enact change, but it is plausible that either differential enforcement of arrest standards or varied degrees of discouragement or encouragement regarding more active policing could explain our observed results.

Our empirical approach does not allow us to discern whether Black and Hispanic LTs are discouraging specific actions, such as reducing certain types of arrests, or if, conversely, white LTs are actively incentivizing effort put towards these marginal arrests. We also cannot take a definitive stance on the normative implications of such results. Arrest declines are not clearly targeted towards arrests of a certain quality (as proxied by court outcomes), non-index arrests may be valuable, and, though we find no significant declines in index arrests, we are not able to definitely rule out quantitatively meaningful changes in these more serious arrests. It is plausible that discouragement (or less encouragement) of more active/aggressive policing generates reductions in potentially both necessary and non-necessary arrests. At the same time, the observed reductions in low-level and discretionary arrests could potentially enhance overall well-being, as many have noted the prevalence of overly aggressive policing and arrests in minority neighborhoods, which is often perceived as harsh and unfair by residents (see US Department of Justice (2017)). A reduction in low-level arrests could also foster improved police-community relations.

We can conclude, however, that leadership matters. We have clear evidence that some LTs are able to influence team-level outcomes. It would be valuable for future work to identify the characteristics, behaviors, and policy and practices that these LTs use to alter arrest totals, composition, and quality. This could then be used to inform interventions targeted at LTs of all races.

6 Conclusion

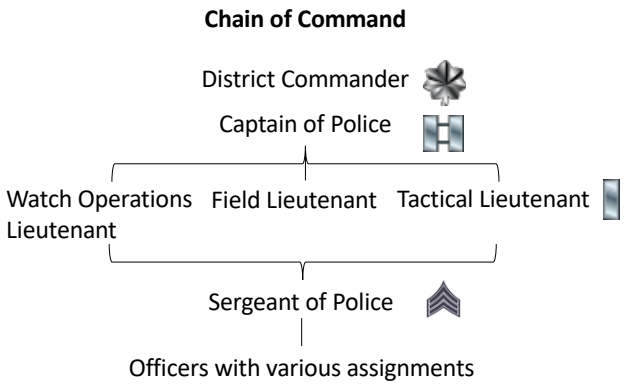
Exploiting day-off group-induced variation in whether or not a Black or Hispanic LTs is predicted to be on duty within a given district and watch, we find robust evidence that arrests decline under the supervision of minority LTs. The observed arrest declines are concentrated among less severe public order arrests and, perhaps, drug-related arrests. We do not find statistically significant declines in index arrests. The quality of arrests, as indicated by court outcomes, does not change. Reductions do not appear to be driven by differential team composition, crime conditions, or officer allocation; nor do they appear to be driven by differences in in-theater supervision by field or tactical LTs or sergeants. Instead, they appear most likely to be the result of differential supervisory practices by watch operations LTs stationed in the district headquarters.

Our results, which document variation in district-watch level outcomes across supervisors of different races, speak to the importance of leadership and supervision by highlighting one, of potentially many, dimension of heterogeneity in team outcomes across police leaders. While we see little evidence of differences across other *observable* LT characteristics, this does not imply that our results are driven by immutable characteristics. On the contrary, we expect Black and Hispanic LTs have, among others, different average management practices/policies, emphases, behavior,

and leadership qualities. Further empirical examination would require data that is not currently collected. Better data on management practices and skills, perhaps obtained through interviews, together with data on officers' perceptions of their supervisor's priorities and expectations could help observers identify the skills and practices that affect team level outcomes and identify potential avenues for future interventions. Collecting such data and using it to think carefully about how to expand and improve the supervisory capability of police leaders of all races, would be valuable.

While recent literature has naturally focused on interventions intended to reduce use of force or discretionary arrests for all officers (e.g Adger et al., 2022; Dube et al., 2023), our work, by emphasizing the influence police LTs have, suggests that interventions targeted at managers could be a valuable complement to these broader approaches. A 2017 U.S. Department of Justice investigation into the CPD similarly emphasizes the need for improved supervision. The report identifies several challenges impeding effective supervision within the organization, such as the excessive administrative focus of sergeants and LTs at the expense of supervisory duties, a reluctance to enforce accountability among subordinates, a shortage of sergeants and LTs, the day-off rotation hindering consistent interactions between officers and their supervisors on every shift, insufficient leadership training, and a lack of repercussions for supervisory shortcomings (US Department of Justice, 2017). The report concludes with suggestions geared towards ensuring that "CPD supervisors have the appropriate tools and information necessary to provide meaningful supervision." Our work also highlights the potential promise of intervening at this level. Interventions targeted at LTs and sergeants, which are much smaller groups than all officers, could end up being highly cost effective.

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

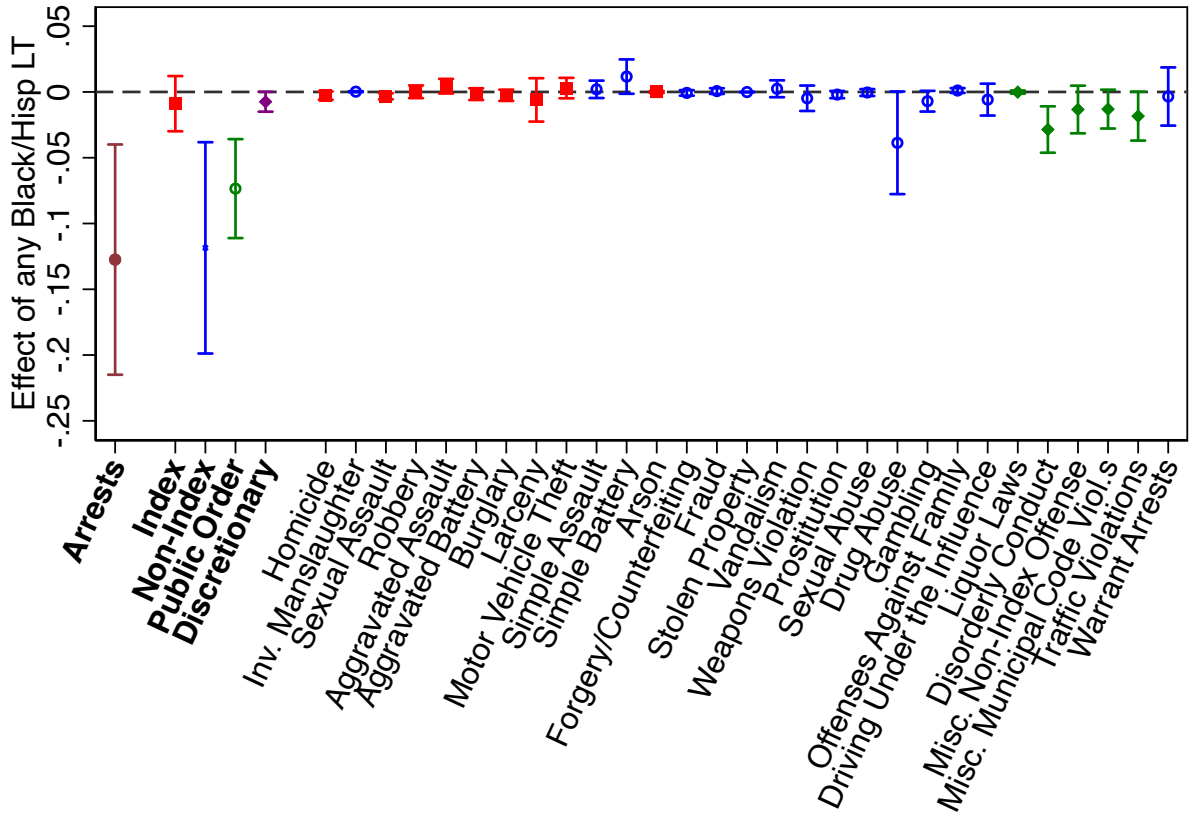
1. Respond to, assume command of, and direct supervisory action for major field incidents occurring in their district
2. Monitor behavior and performance of subordinates
3. In the absence of the watch operations lieutenant, the field or tactical lieutenant will fill the vacancy as directed by district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

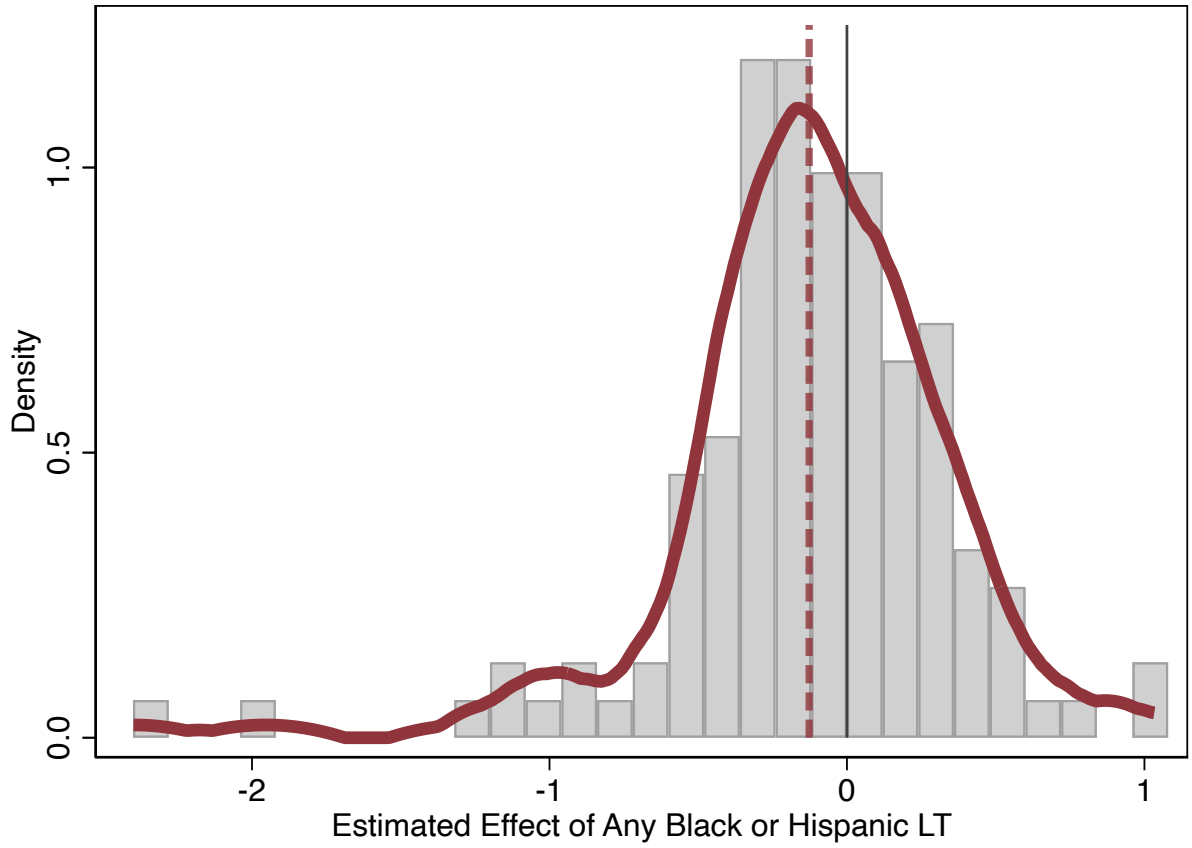
Note: This figure shows the chain of command on the left, splitting out the three types of police lieutenants, and the key duties of lieutenants on the right. Duties are taken from Special Orders S03-03-03 and S03-03-04 (https://directives.chicagopolice.org/#directive/public/6849 and https://directives.chicagopolice.org//#directive/public/6850).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

Figure 2: 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 entry shows the estimated coefficient and 95% confidence interval for overall arrests. The next four entries show coefficients and CIs for arrests in each of our four focal categories (index, non-index, public order, and discretionary). The corresponding point estimates and standard errors for these first five entries can be read from Table 4 (Panel A column (1) and all of Panel B). The subsequent estimates on this figure report results for each separate 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.

Figure 3: 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). 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 -0.125 (median -0.099) arrests per day, very close to the estimated treatment effect of -0.127 from Table 4, which is indicated by the dashed maroon line.

Table 1: Descriptive Statistics

	(1)
Any Pred. Black or Hisp LT	0.20
	[0.40]
Actual:	0.19
	[0.39]
Any Predicted Male LT	0.79
	[0.41]
Actual:	0.72
	[0.45]
Predicted Number of LTs	1.28
	[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: Balance on Crime Conditions

	(1)	(2)	(3)	(4)
			PANEL A: LEAVE OUT ARRESTS	
	All	Same Area	Same watch	Same watch and Area
Any Pred. Black/Hisp. LT	0.0018	-0.0012	0.0003	0.0005
	(0.0023)	(0.0040)	(0.0037)	(0.0069)
Y mean	5.985	4.074	4.834	2.880
			PANEL B: 911 CALLS	
	All 911 Calls (own)	Emergency (own)	911 Calls Leave Out	Emergency Leave Out
Any Pred. Black/Hisp. LT	-0.0031	-0.0054	0.0019	0.0005
	(0.0028)	(0.0060)	(0.0028)	(0.0027)
Y mean	4.594	1.757	8.615	5.939
			PANEL C: PREDICTED ARRESTS	
	Leave-Out Arrests	Leave-Out Arrests (Cubic)	+ 911 Calls	+ 911 and Pred. Officer + Serg. Demo.
Any Pred. Black/Hisp. LT	0.0013	-0.0000	-0.0020	-0.0057
	(0.0063)	(0.0075)	(0.0074)	(0.0087)
Y mean	3.529	3.529	3.529	3.529
Observations	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y

Note: Each entry in this table corresponds to a separate regression of the stated outcome on "Any Predicted Black or Hispanic LT" controlling for what we term the 'baseline FE and controls': Any Predicted Male LT, Predicted number of LTs, and our suite of fixed effects – district×watch×year FEs; day of week FE; and separate FE for each Chicago public holiday. The outcomes in Panel A are leave-out inverse hyperbolic sine of arrests totals from all other districts, from all other districts in the same area, from all other districts in the same watch, and from all other districts in the same watch and area. The outcomes in panel B are the inverse hyperbolic sine of all 911 calls and emergency 911 in the given district-by-watch as well as leave-out 911 calls (i.e. calls in all other districts on that day). Panel C examines predicted arrests, first using only the 4 leave-out arrest variables in Panel A, second adding their squares and cubes, third adding the 911 Call variables from Panel B, and last adding the predicted officer and sergeant demographics. Predictions are formed by regressing district-by-watch arrests on these variables and our baseline fixed effects. The R^2 from the most saturated predictive regression is 0.591 (within-FE is 0.065). *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table 3: Predicted and Actual 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	0.0036	0.0020	-0.0022	0.0003	-0.2918	-0.4312	-0.0025	-0.0060
	(0.0035)	(0.0029)	(0.0040)	(0.0033)	(0.3158)	(0.4659)	(0.0049)	(0.0095)
Y mean	0.184	0.178	0.504	0.796	57.840	41.511	0.212	0.491
				Panel B: Predicted Officer Characteristics				
Any Pred. Black/Hisp. LT	0.0015	0.0047	-0.0025	0.0015	-0.3138	-0.3820	-0.0029	-0.0083
	(0.0038)	(0.0030)	(0.0043)	(0.0036)	(0.2885)	(0.4870)	(0.0052)	(0.0099)
Y mean	0.194	0.180	0.481	0.790	52.360	43.999	0.208	0.507
				Panel C: Predicted Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0267***	-0.0165	-0.0103	-0.0153	0.0254	-0.7658	0.0039	0.0231
	(0.0086)	(0.0103)	(0.0128)	(0.0099)	(0.0694)	(0.5728)	(0.0156)	(0.0230)
Y mean	0.092	0.163	0.718	0.851	5.490	17.862	0.249	0.346
				Panel D: Actual Officer + Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0022	0.0046**	-0.0047	0.0014	-0.0144	-0.5859*	-0.0039	-0.0043
	(0.0020)	(0.0021)	(0.0029)	(0.0023)	(0.2272)	(0.3205)	(0.0038)	(0.0066)
Y mean	0.191	0.167	0.476	0.756	51.358	35.488	0.181	0.432
				Panel E: Actual Officer Characteristics				
Any Pred. Black/Hisp. LT	0.0001	0.0053**	-0.0034	0.0021	-0.0461	-0.6824**	-0.0046	-0.0076
	(0.0023)	(0.0022)	(0.0031)	(0.0026)	(0.2081)	(0.3309)	(0.0040)	(0.0070)
Y mean	0.201	0.168	0.454	0.749	46.567	37.231	0.177	0.442
				Panel F: Actual Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0195***	-0.0025	-0.0160**	-0.0073	0.0130	0.4374	0.0008	0.0264*
	(0.0053)	(0.0073)	(0.0081)	(0.0063)	(0.0332)	(0.4807)	(0.0102)	(0.0153)
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 predicted team compositions. Panels A-C do so for predicted subordinate demographics, based on their day-off groups in January. Panels D-F do so for actual team composition in the district-watch on the given day. The specification is as in Equation 1 replacing the outcome variable with the stated variable. In panels D-F, we set the few missing outcome values that arise equal to 0 and include an extra dummy variable control for missing values. Panels A and D pertain to all officers and sergeants. Panels B and E pertain to just officers. Panels C and F pertain to just sergeants. Columns (6)-(8) examine average arrests, complaints, and use of force in the year prior. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table 4: Reduced Form: Arrests

	(1)	(2)	(3)	(4)
	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.127***	-0.094***	-0.009	-0.020
	(0.045)	(0.034)	(0.009)	(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	-0.009	-0.119***	-0.073***	-0.007*
	(0.011)	(0.041)	(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 fixed effects for district-watch-year, day of the week, and separate dummies for each of the Chicago public holidays. Standard errors, clustered at the district-watch-year, are in parentheses. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

References

ABC News, "George Floyd's Death: As Police Confront Lack of Diversity in Leadership, a Solution Emerges," 2020, *https://abcnews.go.com/US/george-floyds-death-police-confront-lack-diversity-leadership/story?id=71172170.*

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," *The 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.

Alexander, Keith L, Steven Rich, and Hannah Thacker, "The hidden billion-dollar cost of repeated police misconduct," *Washington Post*, 2022.

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

_ , Panka Bencsik, Jesse Bruhn, and Ellora Derenoncourt, "Shots Fired: Crime and Community Engagement with Law Enforcement after High-profile Acts of Police Violence," 2023.

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 and Roman Rivera, "The effect of police oversight on crime and allegations of misconduct: Evidence from chicago," *U of Penn, Inst for Law & Econ Research Paper*, 2019, (19-42).

_ , 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.

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé, "Police officer assignment and neighborhood crime," Technical Report, National Bureau of Economic Research 2021.

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.

_ , Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," *Review of Economic Studies*, 2019, *86* (1), 191–239.

Bloom, Nicholas and John Van Reenen, "Human resource management and productivity," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1697–1767.

Bor, Jacob, Atheendar S Venkataramani, David R Williams, and Alexander C Tsai, "Police killings and their spillover effects on the mental health of black Americans: a population-based,

quasi-experimental study," *The Lancet*, 2018, *392* (10144), 302–310.

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

Capital B News, "Black Police Chiefs, Reform, and Racism," 2022, *https://capitalbnews.org/black-police-chiefs-reform-racism.*

Chattopadhyay, Raghabendra and Esther Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

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," Working Paper 31651, National Bureau of Economic Research September 2023.

Flabbi, Luca, Mario Macis, Andrea Moro, and Fabiano Schivardi, "Do female executives make a difference? the impact of female leadership on gender gaps and firm performance," *Economic Journal*, 2019, *129* (622), 2390–2423.

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

Jordan, Andrew and Taeho Kim, "Strengthening Police Oversight : The Impacts of Misconduct Investigators on Police Officer Behavior," 2023.

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

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.

Lucifora, Claudio, Daria Vigani, and Claudio Lucifora, "Gender Discrimination at the Workplace What If Your Boss Is a Woman ? Work Organization , Work-Life Balance and Gender Discrimination at the Workplace," 2016.

Lum, Cynthia and Daniel S Nagin, "Reinventing american policing," *Crime and justice*, 2017, *46* (1), 339–393.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Quarterly Journal of Eco-*

nomics, 2006, (January).

Matsa, David A. and Amalia R. Miller, "A female style in corporate leadership? Evidence from quotas," *American Economic Journal: Applied Economics*, 2013, *5* (3), 136–169.

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.

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

_ and _, "Does Temporary Affimative Action Produce Persistent Effects? A Study of Black and Female Employment in Law Enforcement," *The Review of Economics and Statistics*, 2019, *94* (4), 1107–1125.

Minni, V, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," *Job Market Paper*, 2023.

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

Owens, Emily and Bocar Ba, "The economics of policing and public safety," *Journal of Economic Perspectives*, 2021, *35* (4), 3–28.

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

Prendergast, Canice, "The limits of bureaucratic efficiency," *Journal of Political Economy*, 2003, *111* (5), 929–958.

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.

Sarsons, Heather, "Interpreting Signals in the Labor Market : Evidence from Medical Referrals," 2017.

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

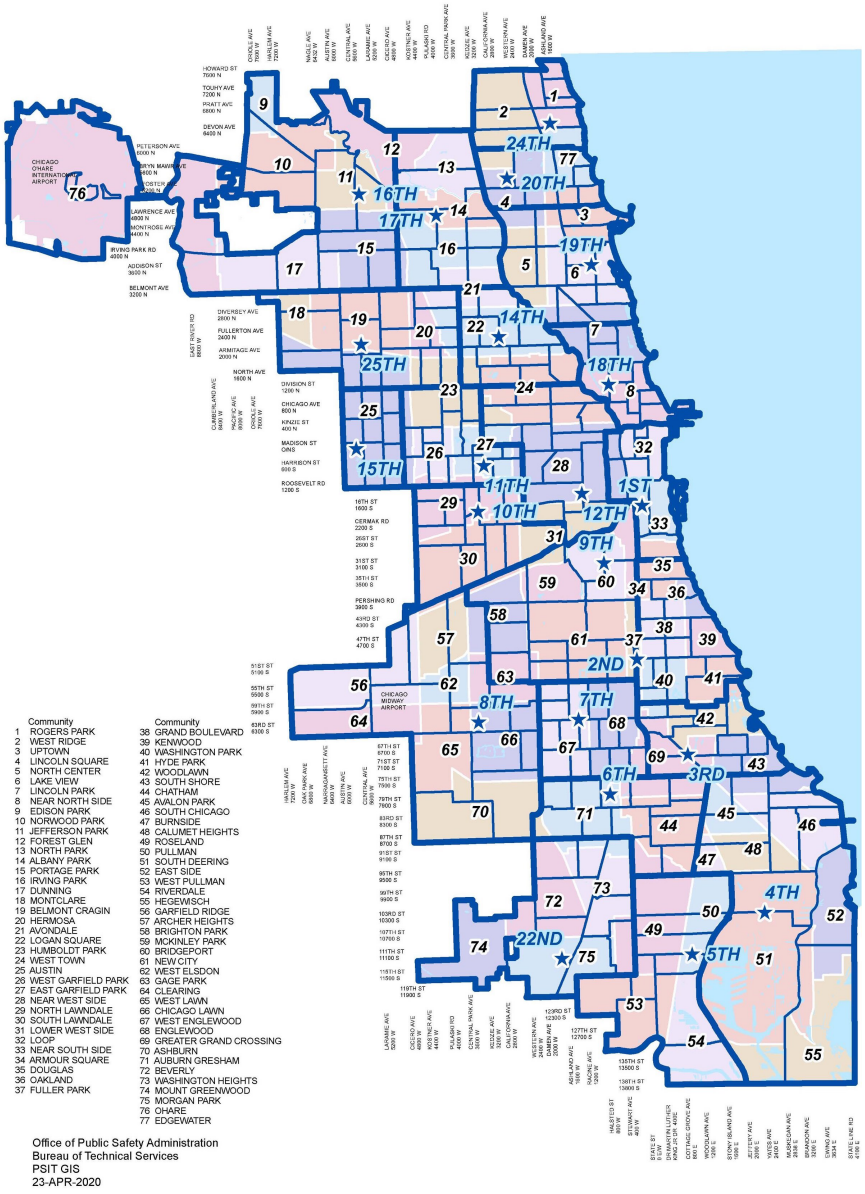
_, "Investigation of the Chicago police department," *US Department of Justice Civil Rights Division and United States Attorney's Office Northern District of Illinois*, 2017.

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.

West, Jeremy, "Racial Bias in Police Investigations," 2018.

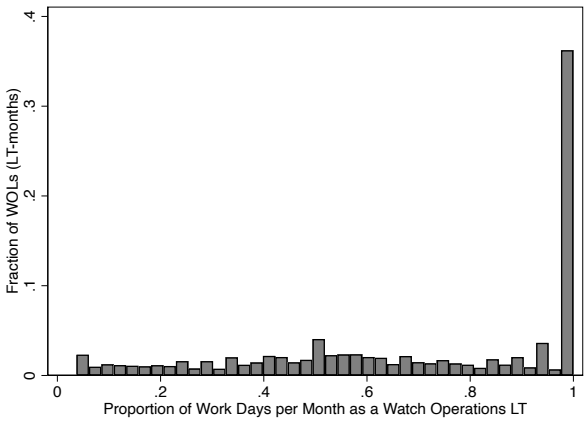
A Online Appendix

Figure A.1: Map of CPD Districts



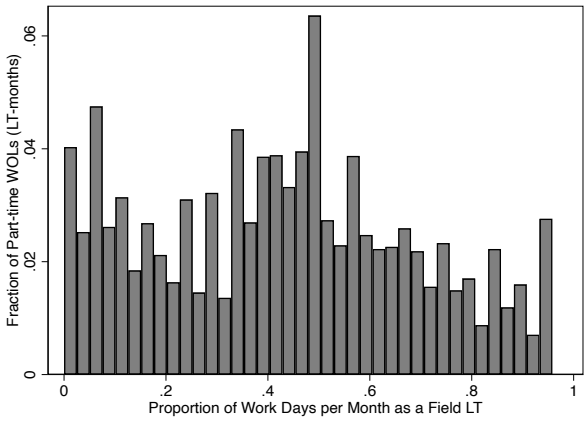
Note: This map, taken directly from https://home.chicagopolice.org/wp-content/uploads/2020/07/CPD-Area-Map.pdf, shows the 22 Chicago police districts (denoted by the solid, thicker blue lines). The district headquarters is denoted by a star. The thinner blue lines denote beats. The multi-colored areas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

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

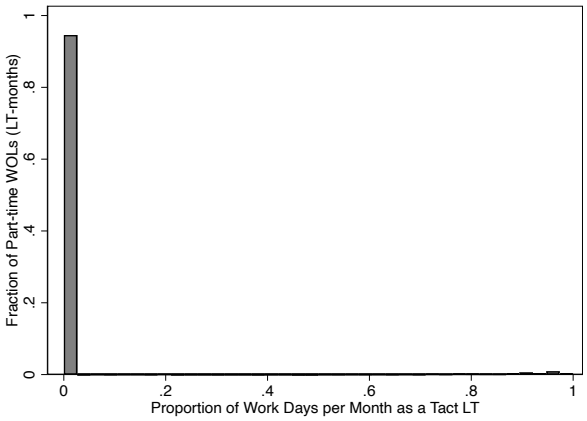


(a) Frac. of Days as WOL

Among Part-time WOLs:



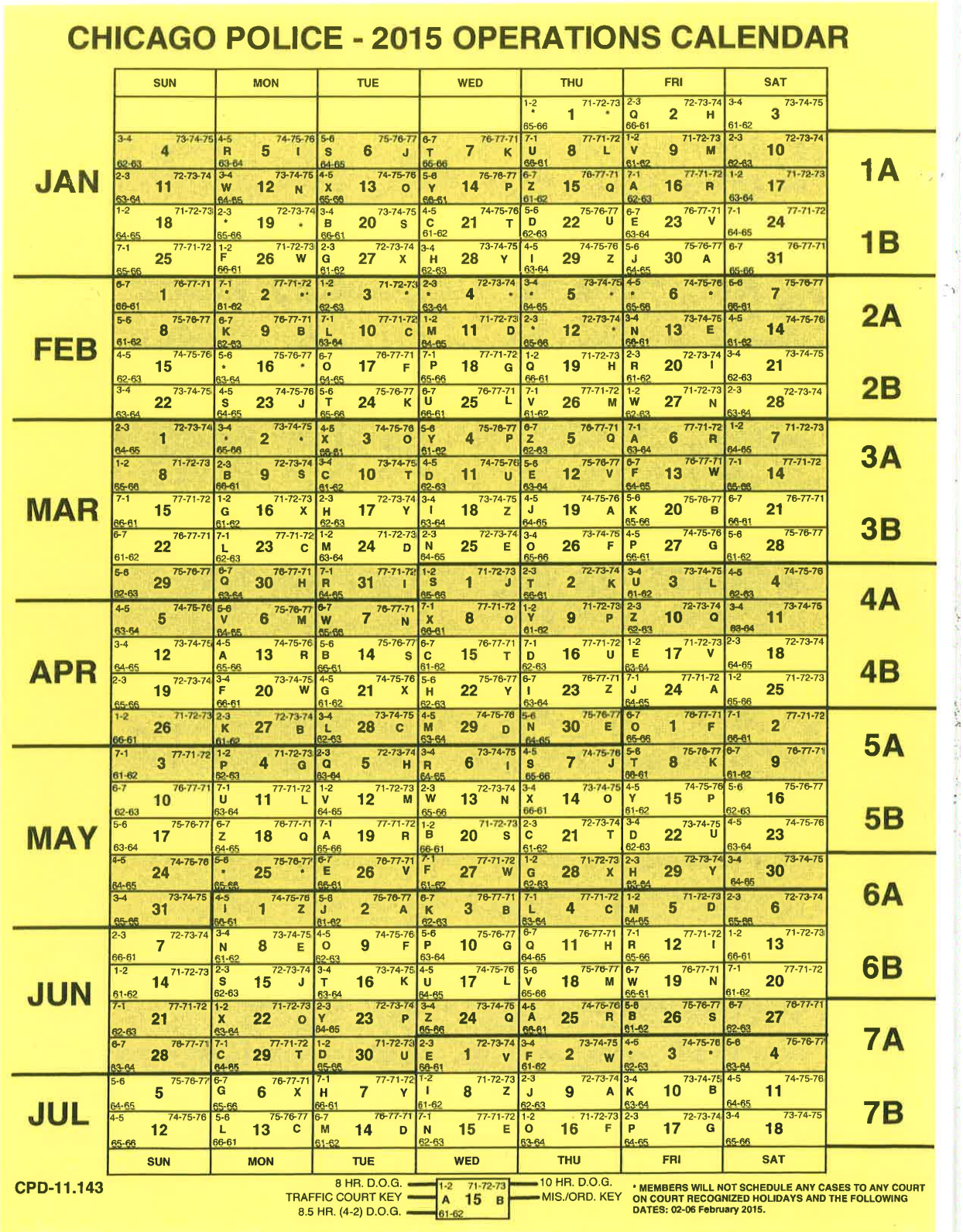
(b) Frac. of Days as field LT



(c) Frac. of Days as Tact LT

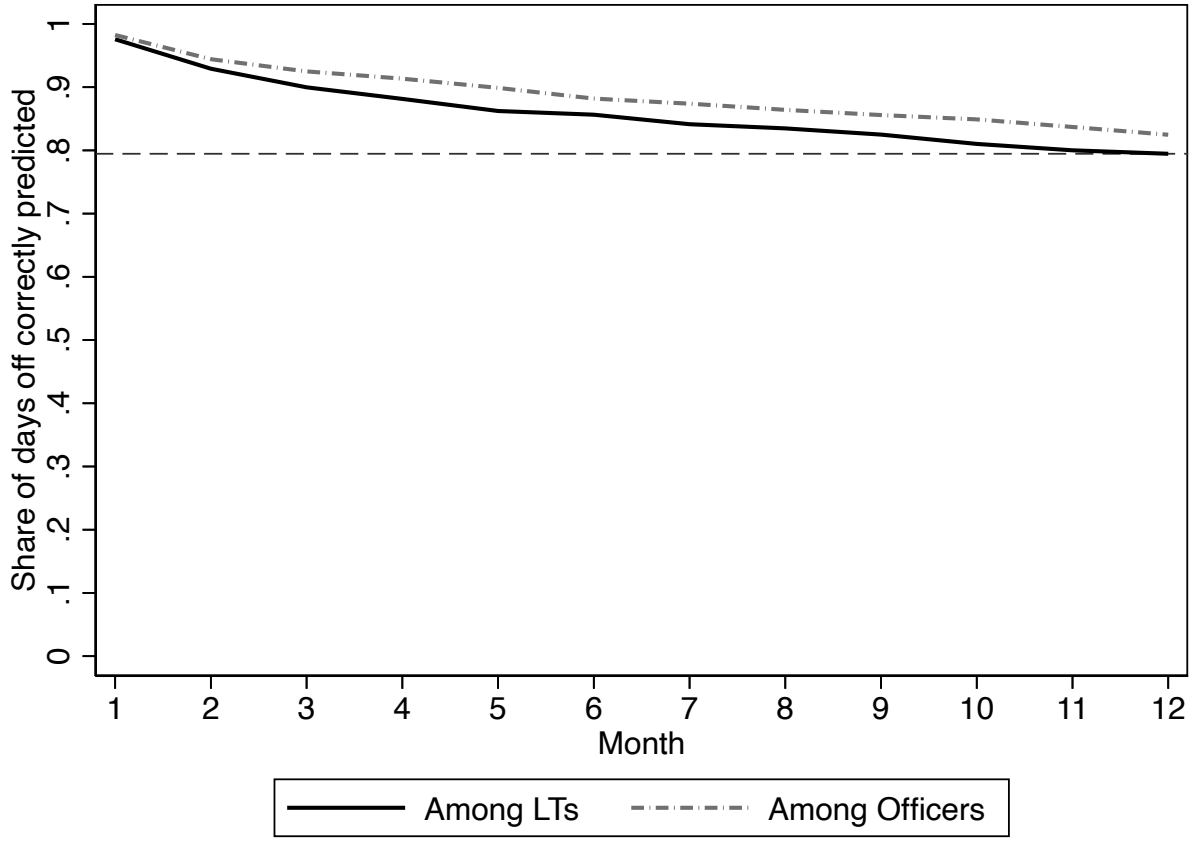
Note: This figure shows how lieutenants who work as watch operations lieutenants split their work days. In the first figure, we show distributions of work days as watch operations lieutenants (WOL) among lieutenants who ever work as WOLs in a month. Specifically, the data are at the lieutenant-month level, and we plot, among those who serve as a WOL at least once in a month, days spent as a WOL in that month. The second and third panels focus on lieutenants 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



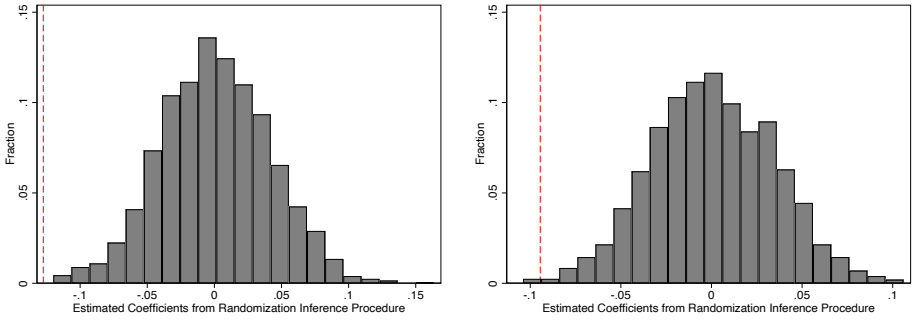
Note: This figure shows the 2015 day-off group calendar obtained from the CPD. The numbers between 61 and 66 in the bottom left corner of each box indicate whether that group has a day off on the given day. For example, on Sunday Jan 4, the 62-63 indicates that day-off groups 62 and 63 are not expected 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: Day-off Group Stability



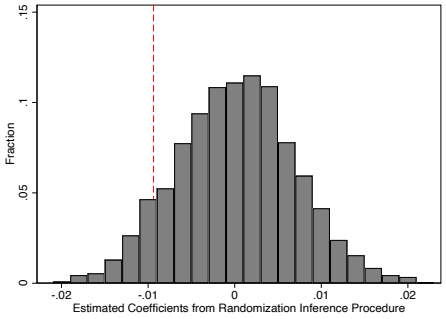
Note: For this figure, we first predict days off for both LTs and officers based only on January attendance records. Then, among the set of days of we predict off, we plot the fraction of such days in each month that we actually observe officers or LTs having a recorded day off, based on attendance records. The solid line shows the fraction of correctly identified days off for LTs, while the dashed line shows this for patrol officers. Overall, the figure shows that day-off groups are relatively stable over time.

Figure A.5: Coefficients From Randomization Inference Procedure

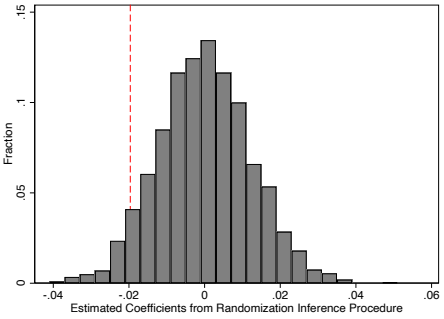


(a) Arrests

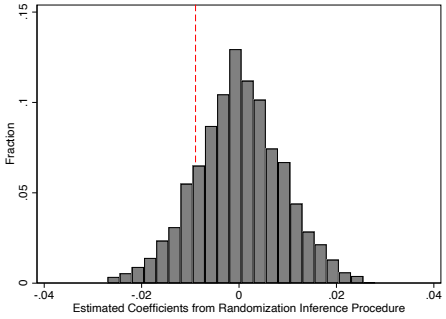
(b) Arrests of Black Persons



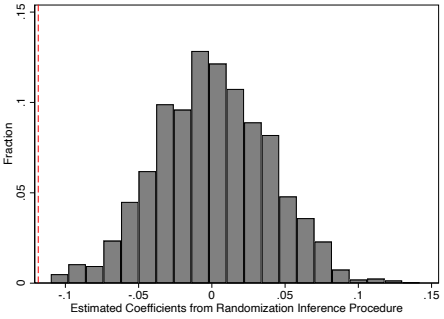
(c) Arrests of White Persons



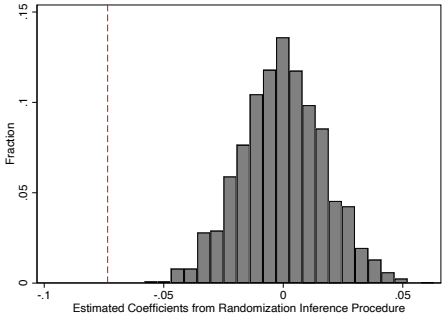
(d) Arrests of Hispanic Persons



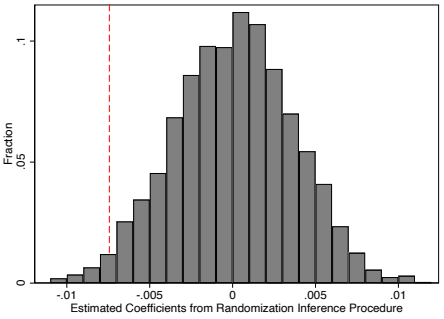
(e) Index Arrests



(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 2000 trials of the randomization inference procedure described in Section 5.2, 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.

Table A.1: Leader Characteristics

	(1) All	(2) Black/Hispanic	(3) White	(4) Difference
Male	0.81	0.73	0.83	0.10
	[0.40]	[0.45]	[0.38]	(0.06)
Tenure	19.50	19.64	19.46	-0.18
	[5.38]	[5.17]	[5.45]	(0.68)
Promoted Since 2004	0.56	0.77	0.51	-0.26***
	[0.50]	[0.43]	[0.50]	(0.06)
Age	45.24	45.87	45.05	-0.82
	[6.03]	[5.04]	[6.28]	(0.69)
Awards	22.79	19.23	23.83	4.59*
	[21.76]	[15.80]	[23.14]	(2.30)
Merit Awards	17.39	13.77	18.45	4.68*
	[20.55]	[14.54]	[21.91]	(2.14)
Complaints	6.33	5.58	6.55	0.97
	[6.78]	[4.99]	[7.22]	(0.72)
N. of Lieutenants	340	77	263	340

Note: This table shows the characteristics of leaders who are active as watch operations or field LTs in our study period. Awards refer to the number of awards that an officer accumulated between 2004 and 2011, and merit awards refer to the number of awards that specifically reward good performance (e.g., Unit Meritorious Award). Complaints are the total number of complaints an officer accrued between 2000 and 2011. "Promoted to LT" is 1 if the leader was promoted to the position of lieutenant between 2004-2020 and 0 if the leader was promoted before that time period. Tenure is defined as the number of years between 2012 and the LTs first CPD appointment date. Standard deviations are in brackets. Column 4 reports a standard t-test for differences in means between columns (2) and (3) with standard errors in parentheses. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A.2: First Stage

	Any Black/Hisp. LT
Any Pred. Black/Hisp. LT	0.456***
	(0.027)
Observations	168,462
Y mean	0.192
Baseline FE and controls	Y

Note: This table shows the results from the same regression as in Table 4 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.3: Balance

	(1)	(2)	(3)	(4)
			PANEL A: LEAVE OUT ARRESTS	
	All	Same Area	Same watch	Same watch and Area
Any Pred. Black/Hisp. LT	0.3024	-0.0686	0.0566	-0.0200
	(0.4544)	(0.1425)	(0.2307)	(0.0799)
Y mean	211.744	33.794	70.581	11.265
			PANEL B: 911 CALLS	
	All 911 Calls (own)	Emergency (own)	911 Calls Leave Out	Emergency Leave Out
Any Pred. Black/Hisp. LT	-0.1789	-0.0247	-2.1561	-0.4071
	(0.1535)	(0.0236)	(4.0021)	(0.4653)
Y mean	64.974	4.143	3898.442	248.550
			PANEL C: PREDICTED ARRESTS	
	Leave-Out Arrests	Leave-Out Arrests (Cubic)	+ 911 Calls	+ 911 and Pred. Officer + Serg. Demo.
Any Pred. Black/Hisp. LT	0.0015	-0.0015	-0.0044	-0.0080
	(0.0072)	(0.0084)	(0.0083)	(0.0093)
Y mean	3.529	3.529	3.529	3.529
Observations	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y

Note: Each entry in this table is analogous to those in Table 2, except instead of applying the inverse hyperbolic sine transformation, all outcome variables are left as raw counts. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A.4: IV Results: Arrests

	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.279***	-0.207***	-0.021	-0.043
	(0.098)	(0.076)	(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	-0.020	-0.260***	-0.161***	-0.016*
	(0.023)	(0.090)	(0.043)	(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 shows the 2SLS results of regressing team-level arrest measures on any Black or Hispanic LT (instrumented for using any predicted Black or Hispanic 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 outlined in Section 3. As usual, all specifications include controls for any predicted male watch or field LT, and the predicted number of watch or field LTs, as well as fixed effects for district-watch-year, day of the week, and separate dummies for each of the Chicago public holidays. Standard errors, clustered at the district-watch-year, are in parentheses. *** p-value $<$ 0.01, ** p-value $<$ 0.05, * p-value $<$ 0.1.

Table A.5: Robustness to Additional Controls

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
				PANEL A: ADDITIONAL CALENDAR TIME FIXED EFFECTS				
Any Pred. Black/Hisp. LT	-0.128***	-0.095***	-0.010	-0.019	-0.009	-0.119***	-0.073***	-0.007*
	(0.045)	(0.035)	(0.009)	(0.015)	(0.011)	(0.041)	(0.019)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
				PANEL B: CONTROLLING FOR LEAVE-OUT ARRESTS AND 911 CALLS				
Any Pred. Black/Hisp. LT	-0.125***	-0.093***	-0.009	-0.019	-0.009	-0.117***	-0.073***	-0.007*
	(0.044)	(0.034)	(0.009)	(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 C: CONTROLLING FOR PREDICTED TEAM COMPOSITION				
Any Pred. Black/Hisp. LT	-0.123***	-0.090***	-0.009	-0.019	-0.011	-0.112***	-0.074***	-0.007*
	(0.041)	(0.032)	(0.009)	(0.014)	(0.011)	(0.038)	(0.018)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
				PANEL D: CONTROLLING FOR EVERYTHING IN PANELS A-C				
Any Pred. Black/Hisp. LT	-0.120***	-0.089***	-0.009	-0.018	-0.011	-0.109***	-0.073***	-0.007*
	(0.041)	(0.032)	(0.009)	(0.014)	(0.011)	(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: This table explores robustness of the results in Table 4 to i) additionally including day-of-the-year (e.g. 32-366) fixed effects and calendar month fixed effects, ii) controlling for the outcome variables in Panels A and B of Table 2, iii) controlling for the outcome variables in Panels A–C of Table 3, and iv) doing i-iii simultaneously.

Table A.6: Controlling for Predicted Officer and Sergeant Demographics, Showing Coefficients

	(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	-0.119***	-0.087***	-0.009	-0.019	-0.008	-0.111***	-0.072***	-0.007*
	(0.042)	(0.033)	(0.009)	(0.014)	(0.011)	(0.038)	(0.018)	(0.004)
Officer and Serg. Share Black	-0.541**	-0.621***	-0.026	0.119*	-0.044	-0.496**	0.034	-0.034
	(0.251)	(0.222)	(0.050)	(0.062)	(0.081)	(0.247)	(0.142)	(0.033)
Officer and Serg. Share Hisp.	-0.476**	-0.538***	-0.003	0.061	-0.089	-0.387*	0.017	-0.020
	(0.241)	(0.202)	(0.052)	(0.084)	(0.058)	(0.229)	(0.131)	(0.023)
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

Note: This table runs the same regressions as in Table 4 but additionally includes controls for the outcome variables in Panel A of Table 3. We omit the share of white officers/sergeants, so that coefficients on the share of Black and share of Hispanic officers and sergeants can be interpreted relative to share non-Black or Hispanic.

Table A.7: Restricting to District-watch-Years with Variation in Any Predicted Black/Hispanic LT

	(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	-0.149***	-0.124***	-0.008	-0.015	-0.014	-0.135***	-0.076***	-0.010**
	(0.055)	(0.044)	(0.009)	(0.017)	(0.012)	(0.051)	(0.024)	(0.004)
Y mean	3.931	2.988	0.302	0.611	0.551	3.379	1.138	0.075
Observations	42115	42115	42115	42115	42115	42115	42115	42115
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to the one in Table 4 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.8: White LTs: Restricting to District-watch-Years with Variation in Any Predicted Black/Hispanic LT

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
Any Pred. White LT	0.157**	0.139**	0.011	0.006	0.014	0.143**	0.080***	0.016***
	(0.063)	(0.057)	(0.011)	(0.016)	(0.013)	(0.059)	(0.022)	(0.005)
Y mean	3.931	2.988	0.302	0.611	0.551	3.379	1.138	0.075
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to the one in Table A.7 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.9: Alternate Inference

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
				PANEL A: CLUSTERING AT DISTRICT-WATCH-YEAR LEVEL (BASELINE)				
Any Pred. Black/Hisp. LT	-0.127***	-0.094***	-0.009	-0.020	-0.009	-0.119***	-0.073***	-0.007*
	(0.045)	(0.034)	(0.009)	(0.014)	(0.011)	(0.041)	(0.019)	(0.004)
				PANEL B: CLUSTERING AT DISTRICT-WATCH LEVEL				
Any Pred. Black/Hisp. LT	-0.127**	-0.094**	-0.009	-0.020	-0.009	-0.119**	-0.073***	-0.007*
	(0.053)	(0.036)	(0.008)	(0.018)	(0.012)	(0.047)	(0.021)	(0.004)
				PANEL C: RANDOMIZATION INFERENCE				
Any Pred. Black/Hisp. LT	-0.127***	-0.094***	-0.009	-0.020	-0.009	-0.119***	-0.073***	-0.007**
	(0.041)	(0.034)	(0.007)	(0.012)	(0.009)	(0.039)	(0.017)	(0.004)
Rand. inference p-value	[0.0015]	[0.0050]	[0.1790]	[0.1085]	[0.3155]	[0.0020]	[0.0000]	[0.0370]
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

Note: This Table reports the same estimates as in Table 4 using different inference procedures. Panel A is as in Table 4. 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 2000 trials; the resulting distributions of estimates (shown in Figure A.5) look normal and centered around 0. The reported standard errors in panel C correspond to the standard deviation of these 2000 estimates, and the p-values correspond to the share of these 2000 trials that produce estimates larger in absolute value than our baseline.

Table A.10: Poisson Maximum Likelihood Regressions

	(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	-0.032***	-0.030***	-0.029	-0.030	-0.015	-0.035***	-0.064***	-0.097*
	(0.011)	(0.011)	(0.029)	(0.023)	(0.019)	(0.011)	(0.016)	(0.050)
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	167793
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
Effect of Any Pred. Black/Hisp LT	-0.111	-0.075	-0.010	-0.020	-0.008	-0.103	-0.064	-0.006
p-value	0.002	0.004	0.308	0.192	0.437	0.002	0.000	0.048

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 4. 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.11: Disaggregating Race of LT

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
				PANEL A: DISAGGREGATING LT RACE				
Any Predicted Black LT	-0.084	-0.063	-0.008	-0.011	-0.015	-0.069	-0.040	-0.005
	(0.083)	(0.069)	(0.012)	(0.018)	(0.017)	(0.078)	(0.029)	(0.005)
Any Predicted Hispanic LT	-0.132***	-0.095**	-0.010	-0.024	-0.005	-0.127***	-0.087***	-0.007
	(0.050)	(0.037)	(0.011)	(0.018)	(0.012)	(0.047)	(0.023)	(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
P-value for Black LT = Hisp LT	0.623	0.692	0.890	0.593	0.609	0.531	0.176	0.801
				PANEL B: DISAGGREGATING WITH CONTROLS				
Any Predicted Black LT	-0.108	-0.081	-0.010	-0.015	-0.019	-0.090	-0.047*	-0.004
	(0.071)	(0.059)	(0.011)	(0.019)	(0.017)	(0.066)	(0.028)	(0.005)
Any Predicted Hispanic LT	-0.113**	-0.079**	-0.008	-0.021	-0.006	-0.107***	-0.084***	-0.007
	(0.044)	(0.033)	(0.011)	(0.017)	(0.012)	(0.041)	(0.022)	(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
P-value for Black LT = Hisp LT	0.961	0.983	0.869	0.781	0.518	0.829	0.272	0.748
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: Instead of examining the effect of any predicted Black or Hispanic LT, we separately examine the effect of any predicted Black LT and any predicted Hispanic LT. All else is as in Table 4. In Panel B, we also add controls for the outcome variables from panels A and B of Table 2 and Panels A-C of Table 3. P-values for the null hypothesis that the coefficient on Any Predicted Black LT is equal to that on Any Predicted Hispanic LT are also reported.

Table A.12: Share Black/Hisp LT

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
				PANEL A: SHARE BLACK/HISP LT				
Share Black/Hisp. LT	-0.182***	-0.140***	-0.018*	-0.021	-0.012	-0.171***	-0.098***	-0.012***
	(0.058)	(0.047)	(0.011)	(0.017)	(0.012)	(0.054)	(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	-0.081*	-0.056	0.000	-0.021	-0.006	-0.075*	-0.052***	-0.003
	(0.046)	(0.034)	(0.009)	(0.017)	(0.013)	(0.040)	(0.020)	(0.004)
$0.5 <$ Share Black/Hisp. LT ≤ 1	-0.185***	-0.141***	-0.020*	-0.020	-0.013	-0.172***	-0.098***	-0.012***
	(0.058)	(0.048)	(0.011)	(0.017)	(0.012)	(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	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 4, 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: Reduced Form: Arrests – White LTs

	(1)	(2)	(3)	(4)
	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. White LT	0.086**	0.062*	0.010	0.013
	(0.041)	(0.036)	(0.007)	(0.012)
Y mean	3.529	2.501	0.326	0.670
	PANEL B: ARREST TYPE			
	Index	Non-index	Public order	Discretionary
Any Pred. White LT	0.002	0.084**	0.058***	0.004
	(0.009)	(0.039)	(0.017)	(0.003)
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 white 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. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1.

Table A.14: Other Outcomes

	Stops	Use of Force	Complaints	Force Complaints
Any Pred. Black/Hisp. LT	-0.1579	-0.0033	0.0006	-0.0003
	(0.1249)	(0.0040)	(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

Note: This table runs the same regression as in Table 4 for 4 additional outcomes: total stops, total use of force incidents, total complaints, and total complaints related to use of force.

Table A.15: Arrests Linked to Court Conviction

	(1)	(2)	(3)	(4)	(5)
			PANEL A: GUILTY ARRESTS		
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.083*	-0.002	-0.081**	-0.048**	-0.076*
	(0.043)	(0.013)	(0.041)	(0.023)	(0.042)
Y mean	2.207	0.440	1.768	0.592	2.150
			PANEL B: NON-GUILTY ARRESTS		
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.082***	-0.005	-0.077***	-0.057***	-0.077***
	(0.029)	(0.007)	(0.030)	(0.021)	(0.028)
Y mean	1.755	0.128	1.627	0.580	1.736
			PANEL C: CHANGE IN GUILTY/NON-GUILTY RATIO FROM THE MEAN		
	All	Index	Non-index	Public Order	Discretionary
Difference in Ratios	0.012	0.117	0.002	0.020	0.012
	[0.663]	[0.586]	[0.936]	[0.720]	[0.664]
Mean Ratio	1.258	3.438	1.087	1.021	1.238
Observations	105,273	105,273	105,273	105,273	105,273
Baseline FE and controls	Y	Y	Y	Y	Y

Note: This table shows the reduced form results of regressing team-level court outcomes on the indicator for any predicted Black or Hispanic LT. We define an arrest to have led to conviction if the case is disposed as either a misdemeanor or a felony. An arrest is defined as non-guilty otherwise (if the case is dismissed, not matched to the court data, or does not have clear case outcomes). In Panel A, we present results from regressions where the dependent variables are arrests linked to conviction outcomes, and in Panel B, the dependent variables are arrests related to non-conviction. Panel C presents tests for how the ratio of conviction-related arrests and non-conviction-related arrests under Black and Hispanic leaders differs from the mean ratio. We add the estimated coefficients for guilty and non-guilty arrests to their respective means and compare the resulting ratio with the ratio of the means.

Table A.16: Arrests Linked to Court Take-up

	(1)	(2)	(3)	(4)	(5)
	PANEL A: COURT TAKE-UP ARRESTS				
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.131**	0.003	-0.134***	-0.080***	-0.120**
	(0.052)	(0.014)	(0.051)	(0.028)	(0.050)
Y mean	2.853	0.488	2.366	0.819	2.784
	PANEL B: NON-TAKEUP ARRESTS				
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.034*	-0.010*	-0.024	-0.026**	-0.034*
	(0.019)	(0.006)	(0.018)	(0.012)	(0.019)
Y mean	1.109	0.080	1.029	0.353	1.102
	PANEL C: CHANGE IN TAKEUP/NON-TAKEUP RATIO FROM THE MEAN				
	All	Index	Non-index	Public Order	Discretionary
Difference in Ratios	-0.040	0.910	-0.078	-0.062	-0.032
	[0.493]	[0.110]	[0.164]	[0.549]	[0.573]
Mean Ratio	2.573	6.100	2.299	2.320	2.526
Observations	105,273	105,273	105,273	105,273	105,273
Baseline FE and controls	Y	Y	Y	Y	Y

Note: This table presents the reduced form results of regressing team-level court take-up outcomes on the indicator for any predicted Black or Hispanic LT. We define an arrest as a take-up if the municipal court considers cases related to the arrest as either a felony or a misdemeanor. An arrest is defined as not being taken up by the court otherwise (if thrown out, not matched to the court data, or has unclear court outcomes). Drug cases are handled in a unique way because drug felonies are not subject to felony review. We consider a felony drug case to be taken up by the court if it reaches a later stage in the process, such as the preliminary hearing or arraignment itself as a felony. In Panel A, we present results from regressions where the dependent variables are arrests linked to court take-up, and in Panel B, the dependent variables are arrests related to non-take-up. Panel C presents tests of how the ratio of the two types of arrests under Black and Hispanic leaders is different from the mean ratio. To conduct these tests, we add the estimated coefficients for take-up and non-take-up arrests to their respective means and compare the resulting ratio with the ratio of the means.

Table A.17: Other LT Characteristics

	(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	-0.125***	-0.093***	-0.009	-0.019	-0.010	-0.114***	-0.074***	-0.006
	(0.047)	(0.035)	(0.009)	(0.015)	(0.011)	(0.043)	(0.020)	(0.004)
Any Pred. Male LT	-0.037	-0.031	-0.003	-0.003	-0.003	-0.034	-0.011	-0.001
	(0.037)	(0.031)	(0.007)	(0.011)	(0.009)	(0.036)	(0.020)	(0.004)
Pred. Num. of LTs	0.037	0.006	0.011**	0.018**	-0.014**	0.051*	0.016	0.003
	(0.028)	(0.022)	(0.005)	(0.009)	(0.006)	(0.027)	(0.016)	(0.003)
Any Pred LT promoted after 2004	-0.004	-0.004	0.000	0.001	0.005	-0.009	0.003	-0.010
	(0.048)	(0.036)	(0.010)	(0.015)	(0.010)	(0.047)	(0.029)	(0.009)
Any Pred LT with tenure $>$ 20	-0.059*	-0.031	-0.012*	-0.016	0.008	-0.067**	-0.006	-0.008
	(0.032)	(0.025)	(0.007)	(0.010)	(0.008)	(0.031)	(0.021)	(0.006)
Any Pred LT with awards $>$ 22	0.021	0.009	0.002	0.009	-0.002	0.022	0.024	0.002
	(0.039)	(0.029)	(0.008)	(0.011)	(0.008)	(0.037)	(0.022)	(0.006)
Any Pred LT with complaints $>$ 6	0.015	0.026	0.001	-0.008	0.006	0.009	-0.014	-0.003
	(0.038)	(0.031)	(0.007)	(0.011)	(0.008)	(0.035)	(0.021)	(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 4, controlling for the other LT characteristics in Table A.1. These are specified in binary format (e.g. any predicted LT promoted since 2004, or any predicted LT with more than 22 awards). We omit Any Pred LT with age $>$ 45 and any Pred LT with merit awards $>$ 17 as they are highly correlated with tenure and total awards, respectively, but main takeaways are stable regardless. See Table A.1 for additional information.

Table A.18: Controlling for Actual Officer and Sergeant Characteristics

	(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	-0.123***	-0.091***	-0.009	-0.020	-0.010	-0.114***	-0.073***	-0.007*
	(0.041)	(0.032)	(0.009)	(0.014)	(0.011)	(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	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 4, controlling for the (potentially endogenous) characteristics of the actual work force (i.e. all the outcome variables in panels D-F of Table 3).

Table A.19: 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	-0.130***	-0.096***	-0.010	-0.020	-0.010	-0.120***	-0.075***	-0.007*
	(0.045)	(0.034)	(0.009)	(0.015)	(0.011)	(0.041)	(0.019)	(0.004)
Share Black/Hisp Sergeants	-0.135**	-0.078	-0.015	-0.044*	-0.030	-0.106*	-0.012	-0.023***
	(0.068)	(0.059)	(0.014)	(0.023)	(0.018)	(0.063)	(0.034)	(0.007)
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 as in Table 4, but adds the share of predicted Black or Hispanic sergeants on duty. In addition to the usual controls, we add controls for the predicted share of male sergeants, and the predicted number of sergeants on duty. The mean share of minority sergeants is 0.26 with a standard deviation of 0.21 and the mean number of sergeants is 5.49.

Table A.20: Disaggregating LT Roles

	(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. Watch Op. LT	-0.184**	-0.126**	-0.033***	-0.019	-0.022	-0.162**	-0.074**	-0.004
	(0.079)	(0.061)	(0.012)	(0.025)	(0.017)	(0.070)	(0.031)	(0.006)
Any Pred. Black/Hisp. Field LT	-0.063	-0.047	0.003	-0.017	0.004	-0.067	-0.063***	-0.004
	(0.044)	(0.035)	(0.010)	(0.014)	(0.012)	(0.041)	(0.023)	(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	0.160	0.240	0.017	0.929	0.182	0.222	0.787	0.997

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 a 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 LTs, any predicted male field LT, and any predicted male watch LT. P-values for a test of equality of the two coefficients are also reported.

Table A.21: Disaggregating LT Roles – White LTs

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
Any Pred. White Watch Op. LT	0.184***	0.113**	0.034***	0.031	0.024*	0.160***	0.088***	0.009*
	(0.066)	(0.051)	(0.011)	(0.020)	(0.014)	(0.059)	(0.027)	(0.005)
Any Pred. White Field LT	0.009	0.012	-0.003	0.001	-0.008	0.016	0.017	-0.002
	(0.039)	(0.032)	(0.007)	(0.013)	(0.010)	(0.037)	(0.019)	(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
P-value for Watch LT = Field LT	0.014	0.062	0.005	0.242	0.062	0.027	0.034	0.082

Note: Here we allow for separate effects of Any Pred. white watch operations LT and Any Pred. white field LT, as opposed to pooling these two groups. We define a 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 LTs, any predicted male field LT, and any predicted male watch LT. P-values for a test of equality of the two coefficients are also reported.

Table A.22: District-by-day-level Analysis Including Tactical LTs

	(1) Arrests	(2) Black Arrests	(3) White Arrests	(4) Hispanic Arrests	(5) Index	(6) Non-index	(7) Public Order	(8) Discretionary
Pred. Black/Hisp. Watch Op. LT	-0.312***	-0.188**	-0.057**	-0.058*	-0.019	-0.293***	-0.159***	0.008
	(0.109)	(0.085)	(0.024)	(0.034)	(0.027)	(0.102)	(0.051)	(0.010)
Pred. Black/Hisp. Field LT	-0.056	-0.046	0.037	-0.047	-0.049*	-0.007	-0.037	0.001
	(0.118)	(0.078)	(0.026)	(0.046)	(0.027)	(0.112)	(0.078)	(0.018)
Pred. Black/Hisp. Tactical LT	0.013	0.017	0.006	-0.003	0.061*	-0.048	0.057	0.007
	(0.167)	(0.129)	(0.031)	(0.053)	(0.035)	(0.165)	(0.105)	(0.020)
Y mean	10.6	7.5	.978	2.01	1.54	9.05	3.04	.208
Observations	56,154	56,154	56,154	56,154	56,154	56,154	56,154	55,299
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
P-val Watch = Field	0.111	0.210	0.011	0.845	0.425	0.063	0.194	0.724
P-val Watch = Tact	0.119	0.199	0.131	0.366	0.073	0.210	0.068	0.959
P-val Tact = Field	0.794	0.739	0.537	0.623	0.051	0.873	0.588	0.865
P-val Joint Test	0.081	0.221	0.019	0.629	0.130	0.068	0.026	0.919

Note: This table runs our baseline specification at the district-by-day level (as opposed to district-by-watch-by-day level), which allows us to reliably predict whether a tactical LT is working on the given day. Due to their rotating, flexible watch schedule, it is difficult to reliably predict their presence at the district-by-watch level. All else is as in Table 4, except we now separately control for any predicted male LT of each type and the predicted number of LTs of each type.

Table A.23: Stops in Various Neighborhoods

	Stops	High Black	High White	High Hispanic
Any Pred. Black/Hisp. LT	-0.1579	-0.0669	-0.0335	-0.0038
	(0.1249)	(0.0952)	(0.0305)	(0.0339)
Observations	168,462	168,462	168,462	168,462
Y mean	11.4984	6.7528	1.5554	1.9473
Baseline FE and controls	Y	Y	Y	Y

Note: This table runs the same regression as in Table 4 for 4 additional outcomes: total stops, total stops in high-Black, high-white, and high-Hispanic beats. High-Black, high-white, and high-Hispanic beats are defined as those with more than 50% of the respective demographic groups.

Table A.24: Arrests Performance and Awards

	Awards from White WOL	Awards from Black/Hispanic WOL
Index Arrests	0.00151***	0.00034***
	(0.00023)	(0.00010)
Non-index Arrests	0.00038***	0.00005***
	(0.00005)	(0.00001)
Observations	8,940,649	8,940,649
Y mean	.000244	.0000476
Relative Contribution of Non-Index to Index Arrests	25.2%	14.7%

Note: This table uses data of daily summary of arrests and awards for each officer from 2013-2020. We regress the total number of daily awards on each officer's total daily count of index arrests and non-index arrests. Each coefficient represents how much each arrest contributes to the number of awards. In the first column, the dependent variable is the total number of awards given by white watch operations lieutenants (where the role is defined on the day the award is given) and in the second column, the dependent variable is the total number of awards given by Black or Hispanic WOLs. We include individual fixed effects, and an indicator for whether or not the officer is in a tactical team, along with the usual fixed effects for district-watch-year, day of the week, and separate holiday dummies. Standard errors are clustered at the team-by-year level. To assess the contribution of an additional non-index arrest towards an officer receiving an award, as compared to index arrests, we divide the coefficient for non-index arrests by the coefficient for index arrests.

Table A.25: Arrests Patterns over Time

	(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	-0.029	-0.002	-0.006	-0.016	0.009	-0.038	-0.012	-0.007
	(0.063)	(0.054)	(0.010)	(0.022)	(0.016)	(0.058)	(0.029)	(0.005)
Months (Since Feb.)	-0.052***	-0.036***	-0.004***	-0.013***	-0.000	-0.052***	-0.023***	-0.001***
	(0.005)	(0.004)	(0.001)	(0.002)	(0.001)	(0.004)	(0.002)	(0.000)
Any Pred. Black/Hisp. LT X Months	-0.020**	-0.018**	-0.001	-0.001	-0.004*	-0.016*	-0.012***	-0.000
	(0.010)	(0.009)	(0.001)	(0.003)	(0.002)	(0.009)	(0.004)	(0.001)
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 interact our baseline indicator for any predicted Black/Hispanic LT with months since February (i.e. = 0 in Feb, 1 in March, up to 10 in December). This shows that our headline, reduced-form results generally appear to strengthen over the course of a calendar year, as officers spend more time under the management of their LT.