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

Matthew Gudgeon, Andrew Jordan, and Taeho Kim

July 2024

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, Michael Makowsky, Ben Pyle, Roman Rivera, CarlyWill Sloan, Brittany Street, Anna Weber, and seminar participants at the NBER Crime Summer Institute, the Society of Labor Economists Annual Meeting, APPAM, SIOE, 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., 2021a; 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, 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 (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). 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 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

[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, and Graham and Makowsky (2024) who study arrests under lame-duck sheriffs, are exceptions.

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.

Instead, we exploit a quasi-experiment generated by the CPD's 4-day on, 2-day off rotation schedule for 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 Hispanic LT is predicted to be on duty relative to days which they are scheduled to have off.[4]

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.

[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]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.

Since subordinate officers share the same set of possible day-off groups as LTs, it is 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.[5] 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 daily, district-by-watch level outcomes, such as total arrests. 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 3.5% relative to the mean) fewer arrests per day.[6] Arrests of persons of different races decline relatively proportionately, but in practice, since Black people are arrested at higher rates, the absolute number of arrests of Black people falls most, accounting for 75% 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 the most serious offenses, we find a statistically insignificant decline of 0.008 arrests per shift, which is 1.6% 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. We also 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 — and find 10% fewer such arrests under Black and Hispanic LTs.[7]

[5]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.

[6]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 6. In our baseline specification, we also control for the predicted number of LTs on duty and day of the week and holiday fixed effects.

[7]We find no clear effects on use of force, or complaints, but results are noisy. We see no clear change in stops, which

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. This is because the exclusion restriction assumption that Black or Hispanic LTs only affect team outcomes on days they actually work is strong. Nevertheless, given we expect leaders might continue to exert some influence in their absence, it is likely that the true average effect of a Black or Hispanic LT lies between the reduced-form and IV estimates (which are approximately twice as large).

We also link arrests to court outcomes in order to explore whether the share of arrests that result in conviction change under Black and Hispanic LTs (as in Mas (2006), Rivera (2022), and Weisburst (2022)). This share can change for two reasons: a change in the distribution of arrest types or a change in the quality of the arrest (i.e. the conviction rate conditional on arrest type). We estimate no clear change in the overall share of arrests that result in conviction. However, consistent with the observed shift in the composition of arrests under Black and Hispanic LTs towards a higher share of index arrests, we do see a statistically significant increase in the share of arrests that result in index arrest convictions. Since the vast majority of convictions stem from non-index arrests (even though these have a lower conviction *rate*), this change is not sufficient to significantly alter the overall share of arrests resulting in convictions. Beyond this composition effect, we see no evidence that Black and Hispanic LTs affect arrest quality — within the various arrest categories the share of arrests that lead to conviction is unchanged.

We subject these headline results to extensive robustness checks, including employing a randomization inference approach and associated placebo check, adding additional controls, estimating district-by-watch-specific effects, using poisson estimation, and specifying the treatment in different ways. 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 team composition. These factors are generally balanced across the relevant sets of days, and controlling for them changes very little. Indeed, when we run a specification at the officer level and include officer fixed effects, so that we estimate how the *same* officer changes

could be viewed as one measure of officer activity, which decline by a statistically insignificant 1.2%.

their arrest behavior on days with and without a predicted Black or Hispanic LT, we continue to detect effects. 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 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 supervisory practices from the district headquarters, as opposed to leaders' actions in the field. Watch operations LTs are responsible for approving arrests and can make it difficult to obtain approval for certain arrests, and, in very rare cases, even refuse to approve an arrest. We see hints that Black and Hispanic LTs might release people brought into the station without charge more often (p-value $= 0.142$), but these events are too rare ($< 1.5\%$ of cases) to directly drive results. While the threat of releasing arrestees without charge could still be important, LTs' preferences for how arrests are made can also filter down through softer forms of leadership or through how they use their other spheres of influence to punish or reward officers. LTs grant coveted overtime or leave requests, take disciplinary actions, and make award nominations. Consistent with a potential preference for focusing more on serious index crimes and less on lower level crimes, we see that Black and Hispanic LTs are more likely to issue overtime for extension of a tour of duty linked to index arrests than white LTs (and less likely to do so for public order arrests). We also find that Black and Hispanic LTs are less likely to make award nominations for commendable performance following lower-severity arrests.

The normative implications of the observed arrest declines depend on the crime reducing benefits and any increased court capacity, distortionary, or legitimacy costs of such marginal arrests (Owens, 2020). In a context where there are significant concerns about overly-aggressive policing and associated harms (US Department of Justice, 2017), and where typical personnel and oversight strategies tend to encourage arrests without factoring in potential legitimacy costs (Owens and Ba, 2021), the observed declines under Black and Hispanic leaders could be welfare-enhancing. On the other hand, in a setting where the marginal benefits of enforcement are higher, such as when legitimacy costs are small and marginal arrests prevent costly and serious crimes in the future, the observed declines in arrests could reduce welfare.

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 supervisors in the police setting, where

incentives and oversight may differ from private firms and where supervisors 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, 2019; 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. Recent work has found that supervisors can matter in specific contexts.[8] Additionally, Kapustin et al. (2022) show that management quality of *executive* officers matters for policing outcomes, documenting that district commanders in the CPD have broad influence over patrol behavior (e.g. stops) and respond causally to a management intervention designed to provide data-driven input into crime prevention strategies. Yet, we know less about whether first line police supervisors (who are not setting crime prevention strategies) impact key outcomes, like arrests. Our findings speak to the importance of these supervisors and their practices. Our findings also suggest that more emphasis could be placed on understanding how police leaders exert influence and at crafting possible interventions targeted at supervisors – which could be a valuable and potentially cost-effective complement to approaches targeted at all officers (e.g Adger et al., 2022; Dube et al., 2023).

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, Section 5 presents key results and robustness, and Section 6 discusses the mechanisms through which LTs exert their influence. Section 7 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).[9] 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

[8]For example, Rim et al. (2022) find that white police supervisors are less inclined to nominate Black officers conditional on work performance than white officers, Frake and Harmon (2023) find that officers working under supervisors with a high complaint history are more likely to receive complaints, and Bindler and Hjalmarsson (2023) show that female police supervisors in Seattle reduce subordinate gender gaps in frisking behavior.

[9]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.

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.[10] Our outcomes are constructed at the daily, district-by-watch (or shift) level (e.g. total arrests on a given day within that district-watch).

Frontline personnel include police officers and sergeants, who play essential roles in the day-to-day operations of law enforcement. At the lowest rank and serving on the front lines of law enforcement, police officers are responsible for various policing duties in the community. Their primary objectives include maintaining order, protecting life and property, responding to crimes and disturbances, conducting initial investigations, and apprehending suspects. Police officers constitute the majority of sworn officers, accounting for approximately 87% of personnel in CPD districts (71% when excluding administrative functions).[11] 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 evaluation (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. 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

[10] 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.

[11] 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.

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 not 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.[12] Given the substantial overlap in the duties performed by watch operations and field LTs and the fact that field LTs can take over for watch operations LTs, we treat these roles interchangeably in our baseline specification. Specifically, we 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 outlined below). We also explore specifications that separate out these 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.

[12]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).

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 six primary 8.5 hour day-off groups (these groups are named 61, 62, ..., 66).[13] 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 January 6th and 7th. 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-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 (which they will begin adhering to in January). The district commanding officers (the watch operations 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

[13] 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.

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. In January of all years we correctly predict 98% of days off for both officers and LTs. From February-December, we correctly predict 87% of days off for officers and 85% of days off for LTs. This match rate declines over time but remains over 79% (see Figure A.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 when most of our data ends.

Our data include demographics, ranks, and the unit assignments of all sworn officers in the CPD from 1946 to 2020. Between 2010 and 2020, we also have daily attendance and 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 have days off. We use these calendars in January of each year paired with assignment and attendance

patterns in that month to infer each officer's day off group and in turn their scheduled work days for the rest of the 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 for 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 and 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 2013-2016 (instead of 2013-2020) because we need follow-up time to observe conviction.

We also have several additional outcomes, all spanning the 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. In Section 6, when we assess how LTs value various types of arrest performances, we utilize additional datasets covering release without charge decisions, overtime approvals, and awards. We discuss these in more detail in that section.

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 a daily, district-by-watch level outcome. The data encompasses 63 district-watch combinations covering 504 district-watch-years.[14] 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 daily arrests, with an average of around 0.5 index arrests and 3 non-index arrests per day.

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 few significant disparities.[15] Black and Hispanic lieutenants are, however, 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

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

[15] 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.

calendar year.[16] 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. 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.[17]

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 operations 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 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.

[16] 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.

[17] 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).

4.2 Constructing Predicted Work Schedules

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 operations 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. This imputation is very accurate. We correctly identify 98% 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 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.

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

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).[18] 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. 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,

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

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.

This strategy identifies 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. Additionally, it is worth noting that this strategy recovers the effect of a Black or Hispanic LT on a particular day, contrasting how officers behave when that particular LT is on duty to how a similar set (or in a later officer level specification with officer fixed effects, the same set) of officers in the same district-watch behave on other days under a different LT. Insofar as Black and Hispanic leaders have effects on other officers that persist into days when they are not directly in charge, we will underestimate them. On the other hand, the rotation-based nature of police scheduling means that officers are constantly working under different leaders, so our estimated effects are particularly relevant to policies that promote and deploy more Black or Hispanic LTs across the force.

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 can potentially influence which day-off group they are in, it is important that either the underlying crime conditions are balanced across day-off groups or, at the least, that minority LTs are no more likely to select a given lower crime day-off group than non-minority LTs. Given the rotating nature of days off, this assumption seems reasonable, but we can also probe it empirically.

Using the specification in Equation 1, we explore whether days with any predicted Black or Hispanic LT look systematically different from other days in terms of expected crime.[20] Panel A in

[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.

Table 2 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, and arrests in only nearby districts in the same watch. Panel B shows that 911 calls in the given district-watch are uncorrelated 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. 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) to the set of predictors. Column 4 additionally includes predicted patrol officer and sergeant demographics (based on their day-off groups). Across specifications, predicted arrests are uncorrelated with whether a Black or Hispanic LT is expected on duty. Overall, this 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 for 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 in panel C, we identify a statistically significant relationship between having a predicted Black or Hispanic LT and the share of sergeants who are Black. We do not observe differences along other non-racial dimensions of sergeant characteristics. Out of the 24 team characteristics we study, the only two that are statistically significant pertain to sergeant race. Since one of the roles of the watch operations LT is to designate a sergeant to serve as the district station supervisor, this suggests that LTs might be handpicking 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. We reach similar conclusions using a regression at the officer-level, which allows us to control for officer fixed effects. Thus, any effects of having a Black or Hispanic LT on team-level outcomes are not likely operating through the makeup of the team.

First stage. Finally, we turn our attention to the effective first stage — specifically, evaluating how often, between February and December, we observe a Black or Hispanic LT on-duty when predicted. Deviations from the predicted schedule can arise because a LT changes their day off group or is no longer in their original district-watch, but also (and importantly) because they take leave on scheduled work days (e.g. vacation), and sometimes work on a day off.

[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, 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.

For the LTs that we identify in January, using attendance records from February-December of each year, we correctly predict whether or not the LT has a scheduled day off on about 86% of days. We further correctly predict whether or not the LT is at work on 70% of the days (February-December of each year). The dropoff from 86% to 70% is due to LTs taking leave on scheduled work days. Finally, on 58% of days we correctly predict whether an LT is at work and in the same unit and watch we assigned them to in January. These numbers do not vary significantly by LT race.

When we aggregate up to the unit-watch level and consider 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), we correctly predict whether any Black or Hispanic LT is working on 67% of days. This corresponds directly to a first-stage coefficient of 0.45 (see Table A.2).[22]

The first stage coefficient of 0.45 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 each Black or Hispanic LTs in these district-watch-years to a randomly selected day off group. 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, violating the exclusion restriction.

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

[22] To understand this, we can cross-tabulate our binary prediction of Black and Hispanic LT presence with reality. 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 67% 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.

arrests by 0.13, which is a 3.5% 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 2.8% 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 75% of the total reduction in arrests.

Instrumental variable results are presented in Table A.4. Scaling the reduced form by the inverse of the 0.45 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, moving forward, we prefer to focus on the more conservative reduced form effect. In reality, given we expect leaders might continue to exert some influence in their absence, it is likely that the true average effect of a Black or Hispanic LT lies between the reduced form and IV estimates (which, given our first stage, can always be approximately obtained by doubling the reduced estimates).

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 reduced-form 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 1.6% of the mean.[23] The corresponding individual FBI codes that comprise index arrests are shown in red and also suggest muted effects. In contrast, non-index arrests decline by a highly significant 0.117, or 3.9% 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 0.009) reduction in disorderly conduct arrests, a 0.01 (4% , p-val 0.009) reduction in miscellaneous non-index offenses, a 0.01 (10%, p-val 0.008) reduction in municipal code violations, a 0.02 (5%, p-val 0.009) 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. 0.064). Together, public order arrests and drug

[23]The 95% confidence interval on index arrests spans a decline of 0.03 daily arrests (5.6% of the mean) to an increase of 0.01 arrests (2.5% of the mean).

arrests account for over 88% (0.111/0.125) of the total, observed reduction in arrests.

We also explore how Black and Hispanic LTs affect 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 0.007 arrests, or 10% of the mean. The officer 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. Of course, the normative implications of the observed arrests declines are complex, something we discuss in more detail in Section 6.

The findings so far suggest that the bundle of characteristics associated with race matters for team outcomes. Do we observe similar differences across other standard, observable LT characteristics? In Table A.5, 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 predicted 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 being relatively comparable across race in Table A.1. Second, we generally do not see any clear effect of having a male predicted LT on duty. Nor do we see differential arrests under LTs with more awards, complaints, or LTs promoted more recently. There is some evidence that non-index arrests are lower under LTs with more experience (tenure).

The fact that the bundle of characteristics associated with race is not well captured by any of these observables is consistent with differences in race being in part due to differences in lived experience that are otherwise difficult to proxy for (at least in this relatively limited set of observables). In 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). We expect lieutenants with different backgrounds to have different management practices, leadership qualities, expectations, and other habits, but these are hard to observe in standard administrative data. In Section 6, we probe this further by examining racial differences in how LTs grant overtime and awards. First, however, we conduct several robustness checks to reinforce our main findings and explore other outcomes.

5.2 Robustness and Specification Checks

We discuss several robustness checks in turn. First, we include additional controls. In Table A.6, 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.7). 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 ($0.18 \cdot -0.576$). 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.8 shows that restricting to the 127 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 127 district-watch-years, Table A.9 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 indeed 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.10, 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 -0.125 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 -0.125, one quarter of district-watch-years have a positive treatment effect larger than 0.125. 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.11.

While we initially considered both Black and Hispanic lieutenants together, Table A.12 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 measurements 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.13 we use the share of LTs who are Black or Hispanic. This exercise yields similar conclusions.

5.3 Other Outcomes: Court Results, Police Activity, Public Safety

We now examine how Black and Hispanic LTs affect other outcomes including eventual court outcomes, police stops, and public safety.

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, we restrict our sample to the years 2013–2016.[24] We are interested in eventual court conviction rates, but cannot *directly* study court conviction rates conditional on arrests because our level of observation is a daily district-watch, and it is not uncommon for a district to go an entire watch with no arrests. Instead, we separately estimate the total number of daily arrests that result in conviction and then study how the ratio of these to total arrests changes.

We begin by studying how the share of arrests that end in conviction changes under Black and Hispanic LTs. Any change here could arise because the distribution of arrests changes (e.g. if the share of index arrests increases and these are more likely to lead to conviction than non-index arrests) or because the quality of arrests changes (i.e. the conviction *rate* conditional on arrest). In Panel A of Table A.14, we find no evidence that Black and Hispanic LTs change the overall conviction rate. Given that the share of arrests composed of index arrests increases under Black and Hispanic LTs, this might come as a surprise. However, the overall conviction rate is a function of both index and non-index convictions, and even though index arrests get convicted at somewhat higher rates (0.77 vs 0.53), the vast majority of overall convictions (over 80%) still come from non-index convictions (given the far higher prevalence of non-index arrests). As a result, any effect on the overall conviction rate is dominated by what happens to non-index arrests.

In Panel B, we clarify this point by estimating separate effects on the share of index arrest convictions (among all arrests) and the share of non-index arrest convictions (again among all arrests). We estimate a 0.008 increase (p-value $= 0.054$, or 7% of the mean) in the share of arrests that result in index arrest convictions. This is driven by change in the distribution of arrest types. Black and Hispanic LTs slightly increase the number of guilty index arrests (0.007, statistically insignificant), but decrease the number of arrests by a highly significant 0.226. Relative to the mean, this increases the *share* of guilty index arrests by 0.008, from 0.108 to 0.116. We see almost no change in the share of non-index arrests that result in conviction.

We next explore whether there is a change in the quality of arrests, by looking for changes in the share of arrests that reach conviction within arrest category. In Panel C, we find little evidence that

[24] 90% of cases in Cook County result in a sentence within 2 years and 95% are sentenced within 2.8 years. Our Court data is as of June 2019. Restricting our sample to prior to December 2016 thus gives us at least 2.5 years to observe how arrests are handled by the court.

Black and Hispanic leaders affect the quality of arrests. In column (1), we note that there is no change in the ratio of guilty index to all index arrests (and there is little change in the number of index arrests in this sample). In column (2), we similarly note no clear change in the ratio of guilty non-index to all non-index arrests (despite clear reductions in the number of non-index arrests). Thus, it appears that Black and Hispanic leaders reduce guilty and non-guilty non-index arrests roughly proportionately. While the point estimates in Panel C are generally positive, indicating increases in quality, they are also typically small (though we are relatively under-powered).

Overall, Black and Hispanic leaders affect court outcomes in a manner consistent with the increase in the share of index arrests that occur under their watch: a larger share of their arrests result in index crime convictions. Nevertheless, this change is not large enough to generate statistically detectable improvements in the overall conviction rate. Nor we do detect any clear impacts, positive or negative, on the quality of arrests within various arrest categories.

Other Outcomes: stops, use of force, and public safety Table A.15 explores other outcomes: total stops, total use of force incidents, and total complaints. Stops, which could perhaps be viewed as a measure of activity, decline by a small and statistically insignificant 1.2% (and a 95% confidence interval spans a relatively tight interval of -3.4% to 0.9% of the mean). We do not detect any statistically significant effects of Black and Hispanic LTs on use of force or complaints, though we are unable to rule out quantitatively meaningful changes. 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.

Did the reductions in arrests lead to an increase in crimes? If the decreased arrests under Black and Hispanic LTs stem from a more relaxed enforcement approach, it is possible perpetrators might commit more crimes on shifts supervised by these leaders (for example if fewer police are observed on patrol). In Table A.16, we analyze crime outcomes. We observe a decline of 0.81% in total crimes, significant at the 10% level, and a less precise decline of 0.88% in terms of index crimes. Some of these declines, however, are likely directly attributed to the fewer arrests made under Black and Hispanic LTs, as arrests can mechanically lead to an increase in reported crimes. To address this concern, we assess the effects on crimes not associated with cleared arrests. We observe a modest decline of 0.49% in such crimes under Black and Hispanic leaders, and a similar decline of 0.87% in index crimes without cleared arrests, neither of which is statistically significant. This supports the notion that reductions in arrests did not result in potential bad actors being uncontained and committing further crime on days when Black and Hispanic LTs managed. Of course, our empirical strategy, with its daily variation, does not allow us to measure any potential cumulative or long-term effects of more Black and Hispanic leaders.

6 Channels behind the declines in Arrest: the influence of watch operations lieutenants

We now probe potential channels behind our main results. We begin by ruling out several possible explanations including changes in team composition and policing strategies. Instead, we suggest that watch operation lieutenants' supervisory standards, expectations, and practices regarding arrests differ by race. We present evidence consistent with this in the form of racial differences in outcomes under LTs sphere of influence: releases without charge, overtime grants, and awards.

6.1 Ruling out other explanations

Differential team composition? As we saw in Table 3, the predicted composition of officers across days when Black and Hispanic LTs are expected to be on duty versus days when they are not appears relatively balanced (with the exception of Black sergeants who are more likely to share a day-off group with minority LTs). Not surprisingly, then, our main findings are highly robust to the inclusion of controls for predicted officer and sergeant composition (as seen in Table A.6). However, since LTs can play a role in adjusting daily schedules as needs arise, it is worth further considering the extent to which differences in outcomes could be generated by differences in *actual* (as opposed to predicted) team composition.

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, barely moves our estimates (see Table A.17).

Another way to probe whether differences in team composition across days in which a Black or Hispanic LT is predicted to be on duty matter is to run our regressions at the officer-level. This allows us to incorporate officer fixed effects and compare arrests for the same officer across days with and without a predicted Black and Hispanic LT.

To do so, we construct a data set where each observation is now at the officer day-by-shift (unit by watch) level. We restrict attention to officers under the supervision of LTs who hold the titles of "Police Officer" or "Sergeant of Police." To avoid double counting, we only count arrests for which the officer was the first arresting officer.[25]

[25] Regular patrol officers are attached to every shift (unit and watch) they are actually observed to be working in the data. Tactical officers require some adjustment since these officers' are assigned day or night shifts that do not

In Table A.18, when running regressions at the officer level without including officer fixed effects, we find that arrests decline by 3.6% (compared to our baseline of 3.5%). When we disaggregate based on arrest categories, we observe reductions that are also very similar to those observed at the team-level: non-index arrests decline by 4.5%, and public-order arrests decline by 7.7%. Next, we include officer fixed effects, such that we now compare arrests made by the *same* officer, across days on which a Black or Hispanic LT is predicted to be working and not. While the estimate of the effect of Black or Hispanic LTs is somewhat attenuated (being about 77% as large for overall arrests and non-index arrests and 94% as large for public order arrests), it is not statistically different from the result without officer fixed effects and remains highly significantly different from zero. These within-officer results make it clear that our main findings are not driven by differential officer composition across days.

Changing crime strategies or patrol behavior? It is also unlikely that LTs affect team-level outcomes through changing crime strategies or altering officer assignment to beats. First, watch operations LTs would not typically alter the beat assignment of officers. Beat officers typically 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 beat officers. Moreover, assignments for the more flexible tactical officers are overseen by tactical LTs or higher ranks, and not watch operations LTs. Last, the formulation of crime prevention strategies occurs at a level above the LTs, specifically at the executive officer and district commander levels (Kapustin et al., 2022). Table A.19 provides partial empirical support for this by examining stop behavior. As already seen, total stops do not change significantly under Black and Hispanic LTs. This table further shows that stops across different subsets of beats are similar under Black/ Hispanic LTs and white LTs. Specifically, stops in high-Black, high-Hispanic, and high-white beats (defined on a district-specific basis based on whether or not the beat is at or above its district's 75th percentile population share for the given demographic) all decline by a statistically insignificant 1-2% of the mean.

6.2 The influence of watch operations LTs

Instead, we believe our results are best explained by watch operations LTs of different races differently incentivizing arresting officer behavior. Here, we first show that our affects are driven by watch operations LTs (as opposed field or tactical LTs) and then show that LTs of different races use their levers of influence (e.g. overtime and award grants) differently.

perfectly fit into the standard three-watch schedules of district operations and that alternate over the course of the year. To address this, we assign each tactical team officer's work time to be one of the three watches that overlap the most with the day's modal start and end times within that officer's sub-team.

Effects are concentrated on watch operations LTs. Our baseline analysis combines the effects of watch operation and field LTs. 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 $= 0.15$). 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.[26]

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 operations LT is equal to that for tactical LTs is 0.140, and the p-value on a joint test of equality across all three coefficients is 0.086. 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? Watch operations LTs hold significant spheres of influence over officers. They manage scheduling, which includes granting or denying overtime and vacation requests, and can arrange for officers to change shifts or day-off groups. They grant awards for commendable arrests, conduct complaint investigations, and take disciplinary action when performance deficiencies occur. Importantly, they also ultimately ensure compliance with arrest policies and procedures and have to review and approve all arrests.

In extremely rare cases, LTs might not approve arrests brought into the station at all (in which case the event would not show up in our data as an arrest). We issued a FOIA request for recorded events of 'release without charge,' which indeed are very rare: on average we see 0.053 releases without

[26] 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 operations LTs play a predominant role.

charge per day, relative to 3.529 non-released arrests (or less than 1.5% of cases). Given how rare these events are, it is not particularly surprising that in Table A.23 we cannot detect differences in either the number or the share of releases without charge across days when Black and Hispanic LTs are predicted to work. When we investigate how Black and Hispanic LTs change the share of releases without charge for different types of arrests, some of magnitudes and associated p-values (e.g. $p = 0.142$ for public order related releases) suggest that this (stark) lever may be used more often by Black and Hispanic LTs. Of course, it is also possible that the threat of using this lever is sufficient to disincentivize arrests in the field that the LT might not want to support.

Given how rare these events are, however, we think it is likely that LT preferences also filter down through softer forms of leadership (e.g. instructions, emphasis on standards, etc) or through other spheres of influence. For example, by granting coveted overtime or awards for marginal non-index arrests, an LT could effectively issue an endorsement of more active policing. By threatening to void certain types of arrests, making it burdensome to receive a sign off on an arrest, or not issuing overtime for them, LTs could similarly disincentivize certain arrest types.[27] More generally, through actions like granting or not granting leave and shift-related requests, LTs have tools to reward or punish officers that they like or dislike. To assess whether LTs use these other spheres of influence differently, we next investigate whether Black and Hispanic LTs differ in their overtime grants or use of awards.

Because of potential misuse of overtime, strict guidelines require prior watch operations LT approval before any overtime is worked, and LTs are required to approve only work deemed necessary for overtime. Hence, examining overtime approvals for an extension of an officer's tour of duty (which are often linked to needing to finish up an arrest) and how these differ by arrest type, can provide insight into the types of arrests that different LTs value. In Table A.24, we first show that there is no statistically significant difference in the number of officers who work overtime for an extension of their tour of duty across LTs of different races. However, when we then look at the type of arrests made in connection to the overtime work (i.e. arrests made four hours before until the end of overtime), we note differences across LTs. Black and Hispanic LTs appear significantly more likely to grant overtime requests linked to index arrests and less likely to grant overtime requests linked to public order arrests. These differences are consistent with a differential emphasis

[27] Both current and former CPD officers have shared experiences where the preferences and attitudes of watch operations LTs have filtered down to officers' decision-making processes. When arrest reports fail to meet a watch operations LT's standard, the LT may reprimand the arresting officer and require them to revise the report. Additionally, several public investigations by the Civilian Office of Police Accountability (COPA) show that LTs can face consequences for approving arrests that do not meet the standard of probable cause (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 in arrest approval and hint at how variation in LT standards could lead to disparate arrest outcomes.

on index relative to non-index (and particularly public order) arrests. Furthermore, since overtime is generally sought after due to the pay bump, these differences would likely be meaningful to the officers under their charge.

Last we investigate how Black and Hispanic LTs differ in their award nominations for a given level of performance. While many awards are given out by sergeants, LTs play an important role in nominating awards in recognition for outstanding performance. Although such awards are infrequently granted, exploring how the propensity to nominate officers for awards scales with different arrest types provides insights into the level of importance LTs attribute to these specific arrest categories. We use award data from 2013 to 2020, focusing on award nominations linked to incidents that occurred on a specific day (as opposed to over multiple days) and were specifically given out by LTs. The vast majority of these award nominations are honorable mentions or commendations linked to performance during an arrest or crime incident.

In Table A.25, we first find in column (1) that Black and Hispanic LTs are no less likely to nominate awards overall, despite the reduction in arrests they generate. We then ask how the relationship between daily, team-level arrests and LT award nominations differs on days in which a Black or Hispanic LT is predicted to be working. In column (2), for example, we see that on average each daily arrest increases the number of award nominations by 0.005. This slope is similar on days on which Black and Hispanic LTs are predicted relative to days they are not. In column (3), we see that each index arrest increases the number of award nominations by 0.016. Again, this is similar across LTs of different races. However, in the remaining columns we note that there is a weak positive link between index, public order, or discretionary arrests and award nominations under white LTs. Meanwhile, Black and Hispanic LTs are *less* likely to issue award nominations on days with high public order arrests or discretionary arrests (with p-values on the difference relative to white LTs of 0.083 and 0.053). These results indicate that while Black and Hispanic LTs do not differ from white LTs in their appreciation of index arrest performance, they tend to assign less importance to enforcement activities involving lower-level arrests.

6.3 Discussion

What can and can't we conclude from our 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 award low level arrests relative to their white peers, issue overtime for extended duty tours differentially more for serious index arrests and differentially less for public order arrests, and may be releasing without charge more often. As a result, we suspect that either differential enforcement of arrest standards or varied degrees of discouragement or encouragement regarding more active policing could explain our results. We note that our empirical approach does not allow us to discern whether Black and Hispanic LTs are discouraging specific actions, such as making certain types of arrests, or if, conversely, white LTs are actively incentivizing effort put towards these marginal arrests.

The normative implications of our findings ultimately depend on whether reductions in the marginal non-index arrest relative to the status quo is optimal.[28] In a Becker-style framework, the optimal amount of force or arrest effort in each encounter is the amount that balances the crime reducing benefits with the court capacity, distortionary costs, and legitimacy costs (Owens, 2020, see for e.g.). If the status quo under the CPD over our time period involved too many non-index arrests (i.e. involved arrests that generated lower crime reduction benefits than they increased legitimacy or other costs), then the observed reductions in low-level and discretionary arrests under Black and Hispanic LTs could potentially have enhanced well-being. Consistent with this being a possibility, our data span a period over which the department of justice civil rights division found that the CPD engaged in a pattern of use of force in violation of the Constitution, with minority neighborhoods bearing the brunt of its impacts (US Department of Justice, 2017). The same report also documents how policing in minority neighborhoods is often perceived as harsh and unfair by its residents, and how victims of crimes and their families feel officers do not put sufficient effort into investigating more significant crimes. An excess of arrests is also consistent with typical compensation and oversight strategies that reward arrests without regard for its potential legitimacy costs (Owens and Ba, 2021). Conversely, in a context where the marginal benefits of enforcement are substantially higher, such as when legitimacy costs are low and marginal arrests help prevent costly and serious future crimes, the observed declines in arrests could be welfare-reducing.

We can conclude more definitely, however, that supervisors matters. We have clear evidence that some LTs are able to influence team-level outcomes. Their practices also differ with respect to awards and overtime grants. It would be valuable for future work to identify the other 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.

[28] While we find no significant changes in the number of index arrests or the number of index arrests that eventually result in convictions, the fact that we are not powered to rule out index arrest declines of 5.6% of the mean, which might have higher marginal crime reducing benefits, further complicates welfare statements.

7 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. Instead, they appear most likely to be the result of differential supervisory practices and emphases 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 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, emphases, and other qualities. And, indeed, we see evidence reflective of this in overtime decisions and award nominations. 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.

While recent studies have focused on interventions intended to reduce use of force or discretionary arrests for all officers (e.g Adger et al., 2022; Dube et al., 2023), our work, suggests that interventions targeted at supervisors 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 execssive administrative duties, a reluctance to enforce accountability among subordinates, the day-off rotation hindering consistent interactions between officers and their supervisors, insufficient leadership training, and a lack of repercussions for supervisory shortcomings (US Department of Justice, 2017). The report concludes with suggestions geared towards improving the ability of LTs and sergeants to provide better supervision. Our work similarly highlights the potential promise of intervening at this level, which given it comprises fewer people, could also end up being highly cost effective.

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, 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.

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

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.*

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.

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.

Graham, Shannon R and Michael D Makowsky, "Lame Duck Law Enforcement," *Economics Letters*, 2024.

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

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

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

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

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

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

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

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

Miller, Amalia R and Carmit Segal, "Does Temporary 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, "The economics of policing," *Handbook of labor, human resources and population economics*, 2020, pp. 1–30.

_ **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.

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.

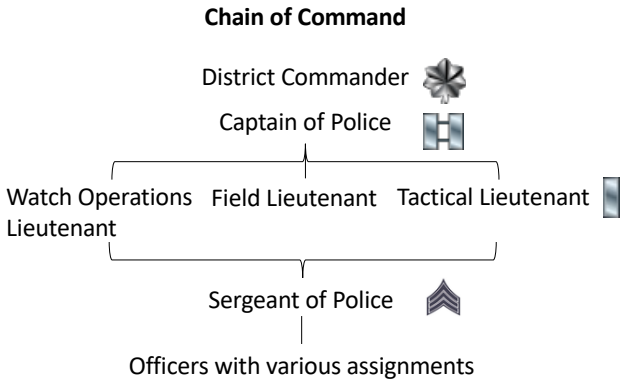
US Department of Justice, "Advancing Diversity in Law Enforcement Report (October 2016)," *US Department of Justice Civil Rights Division*, 2016.

_ , "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.

Figures

Figure 1: Chain of Command and Police Lieutenant Duties



Select Duties

Watch Operations LTs (staffed at all times):

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

Field LTs

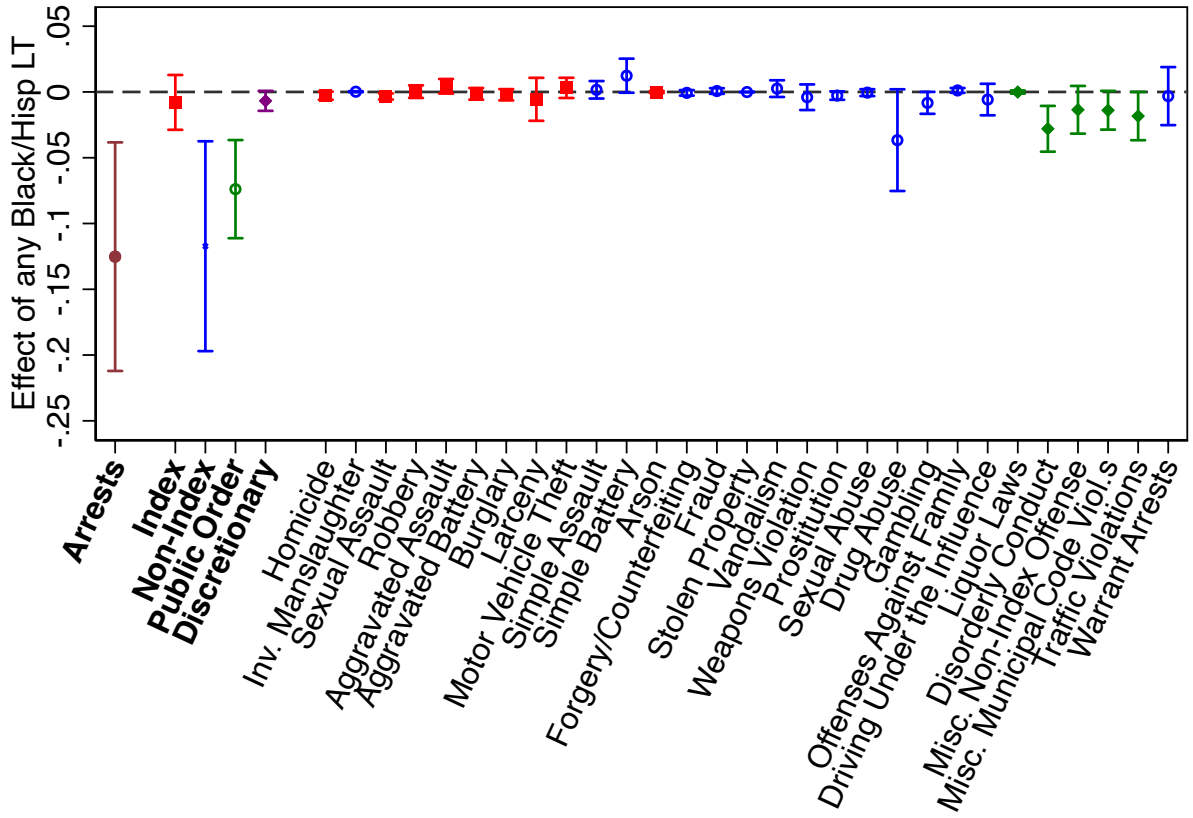
1. Respond to, assume command of, and direct supervisory action for major field incidents occurring in their district
2. Monitor behavior and performance of subordinates
3. In the absence of the watch operations lieutenant, the field or tactical lieutenant will fill the vacancy as directed by district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

Tactical LTs

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

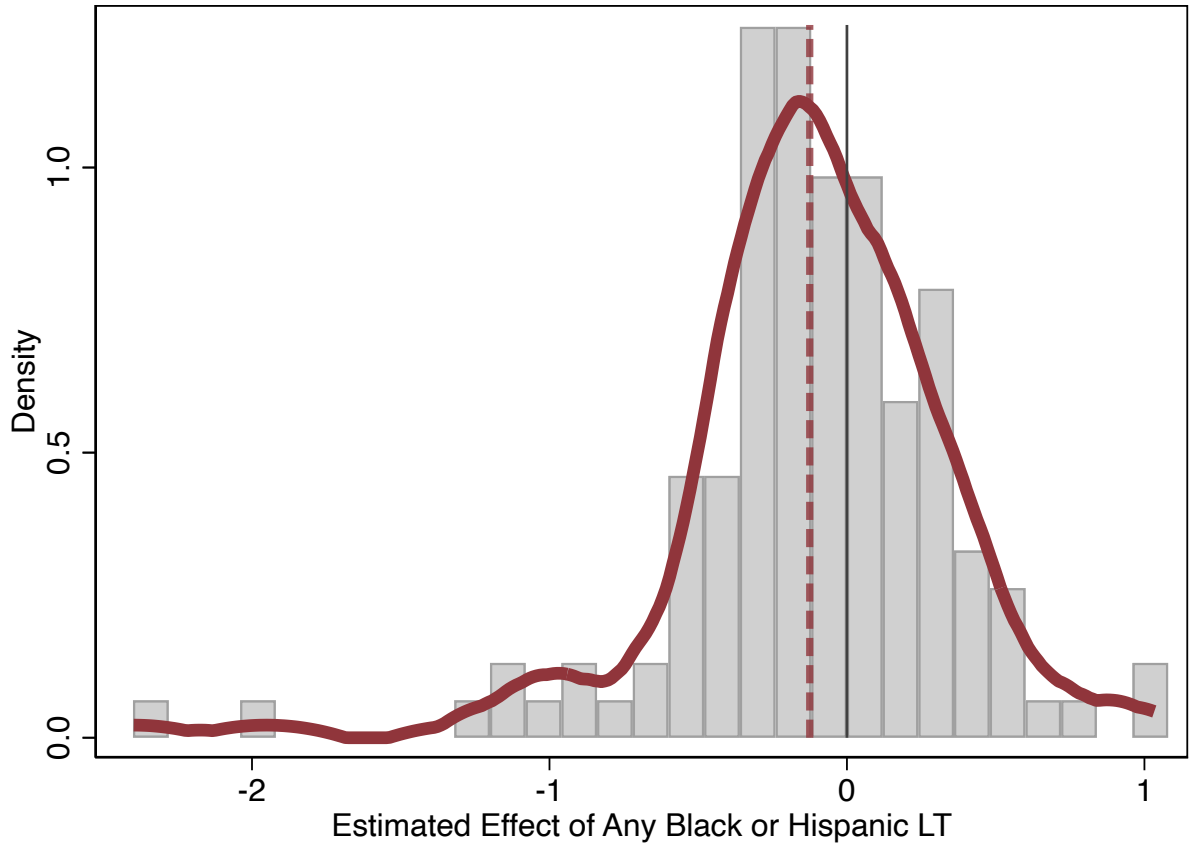
Note: This figure shows the chain of command on the left, splitting out the three types of police lieutenants, and the key duties of lieutenants on the right. Duties are taken from Special Orders S03-03-03 and S03-03-04 (https://directives.chicagopolice.org/#directive/public/6849 and https://directives.chicagopolice.org//#directive/public/6850). 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 -0.123 (median -0.11) arrests per day, very close to the estimated treatment effect of -0.125 from Table 4, which is indicated by the dashed maroon line.

Tables

Table 1: Descriptive Statistics

	(1)
Any Pred. Black or Hisp LT	0.20
	[0.40]
Actual:	0.19
	[0.39]
Any Predicted Male LT	0.79
	[0.40]
Actual:	0.72
	[0.45]
Predicted Number of LTs	1.28
	[0.66]
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	0.0020	-0.0011	0.0005	0.0007
	(0.0023)	(0.0040)	(0.0037)	(0.0068)
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	-0.0030	-0.0051	0.0022	0.0007
	(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	0.0018	0.0010	-0.0011	-0.0034
	(0.0063)	(0.0076)	(0.0074)	(0.0085)
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	0.0040	0.0017	-0.0023	0.0005	-0.1882	-0.4324	-0.0021	-0.0060
	(0.0034)	(0.0029)	(0.0040)	(0.0032)	(0.3145)	(0.4633)	(0.0048)	(0.0095)
Y mean	0.184	0.178	0.504	0.796	57.922	41.494	0.212	0.491
				Panel B: Predicted Officer Characteristics				
Any Pred. Black/Hisp. LT	0.0017	0.0045	-0.0025	0.0018	-0.1860	-0.4042	-0.0024	-0.0085
	(0.0038)	(0.0030)	(0.0043)	(0.0035)	(0.2874)	(0.4833)	(0.0052)	(0.0100)
Y mean	0.194	0.180	0.481	0.790	52.420	43.994	0.208	0.507
				Panel C: Predicted Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0265***	-0.0171*	-0.0089	-0.0150	0.0036	-0.7281	0.0025	0.0223
	(0.0084)	(0.0102)	(0.0126)	(0.0098)	(0.0672)	(0.5691)	(0.0155)	(0.0228)
Y mean	0.092	0.163	0.718	0.851	5.508	17.839	0.249	0.346
				Panel D: Actual Officer + Sergeant Characteristics				
Any Pred. Black/Hisp. LT	0.0022	0.0045**	-0.0046	0.0014	-0.0100	-0.6658**	-0.0039	-0.0040
	(0.0020)	(0.0021)	(0.0028)	(0.0023)	(0.2247)	(0.3318)	(0.0037)	(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	0.0052**	-0.0033	0.0020	-0.0461	-0.7699**	-0.0044	-0.0072
	(0.0022)	(0.0022)	(0.0031)	(0.0026)	(0.2058)	(0.3444)	(0.0039)	(0.0069)
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***	-0.0026	-0.0154*	-0.0072	0.0174	0.4679	-0.0009	0.0264*
	(0.0052)	(0.0072)	(0.0081)	(0.0063)	(0.0331)	(0.4777)	(0.0102)	(0.0151)
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

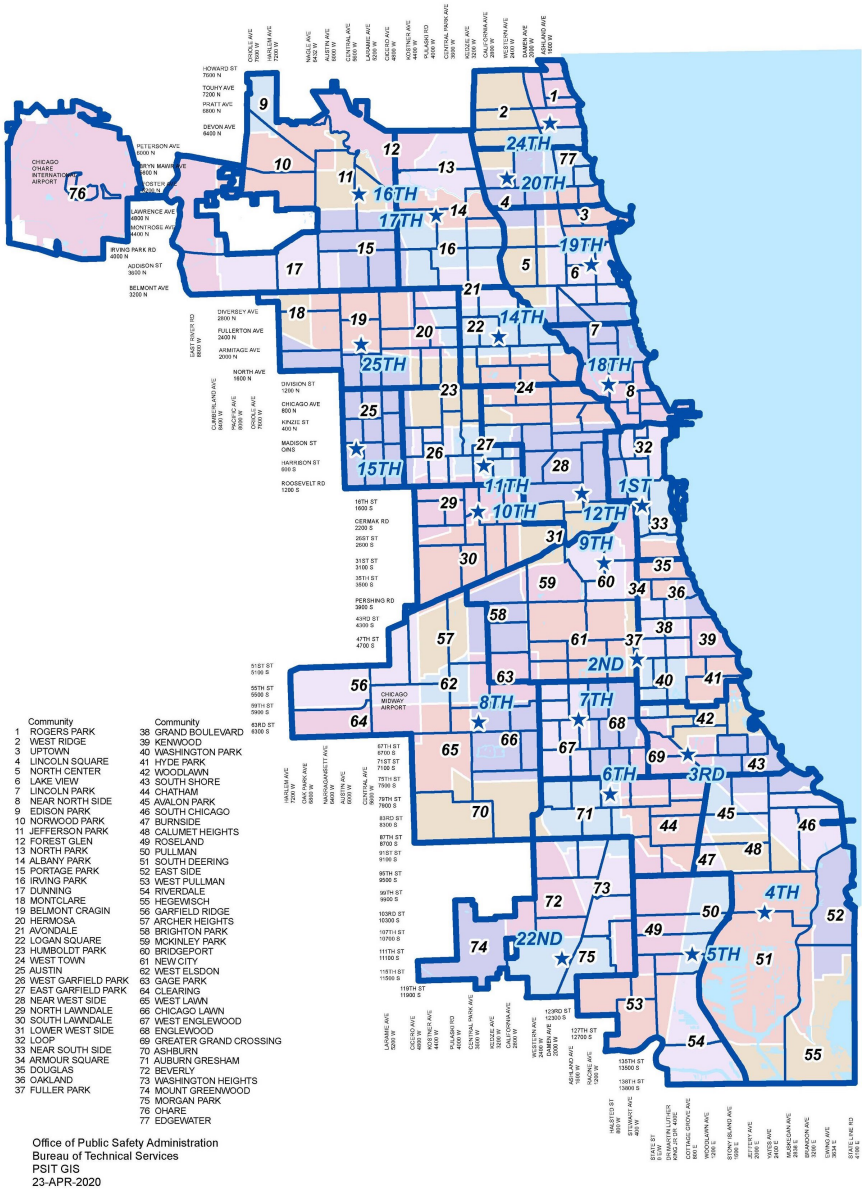
	(1)	(2)	(3)	(4)
	PANEL A: ARRESTS			
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.125***	-0.094***	-0.009	-0.019
	(0.044)	(0.034)	(0.009)	(0.014)
Y mean	3.529	2.501	0.326	0.670
	PANEL B: ARREST TYPE			
	Index	Non-index	Public order	Discretionary
Any Pred. Black/Hisp. LT	-0.008	-0.117***	-0.074***	-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.

A Online Appendix

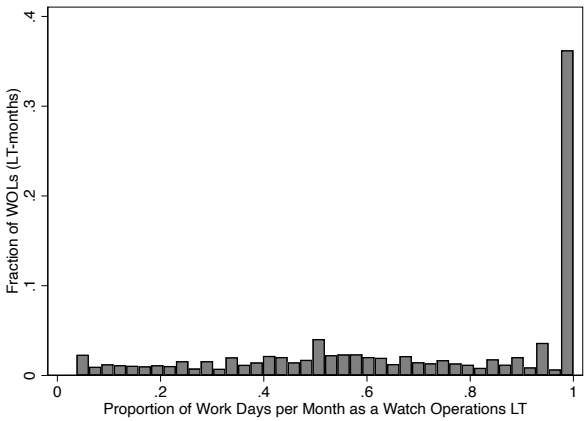
Appendix Figures

Figure A.1: Map of CPD Districts



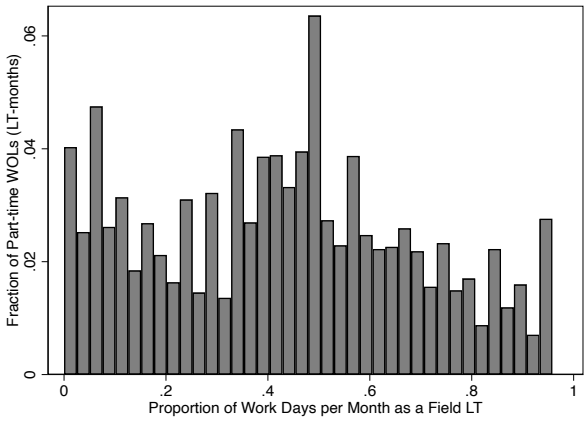
Note: This map, taken directly from https://home.chicagopolice.org/wp-content/uploads/2020/07/CPD-Area-Map.pdf, shows the 22 Chicago police districts (denoted by the solid, thicker blue lines). The district headquarters is denoted by a star. The thinner blue lines denote beats. The multi-colored areas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

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

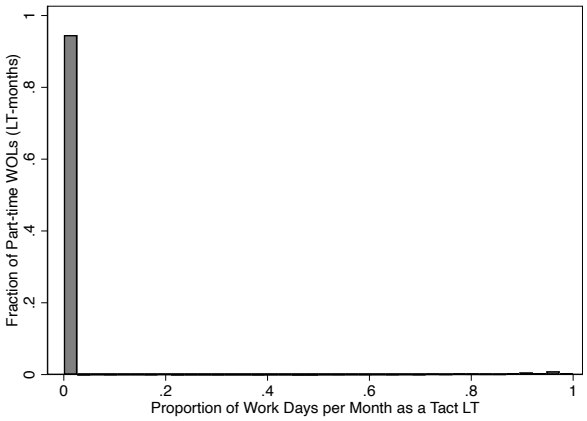


(a) Frac. of Days as WOL

Among Part-time WOLs:



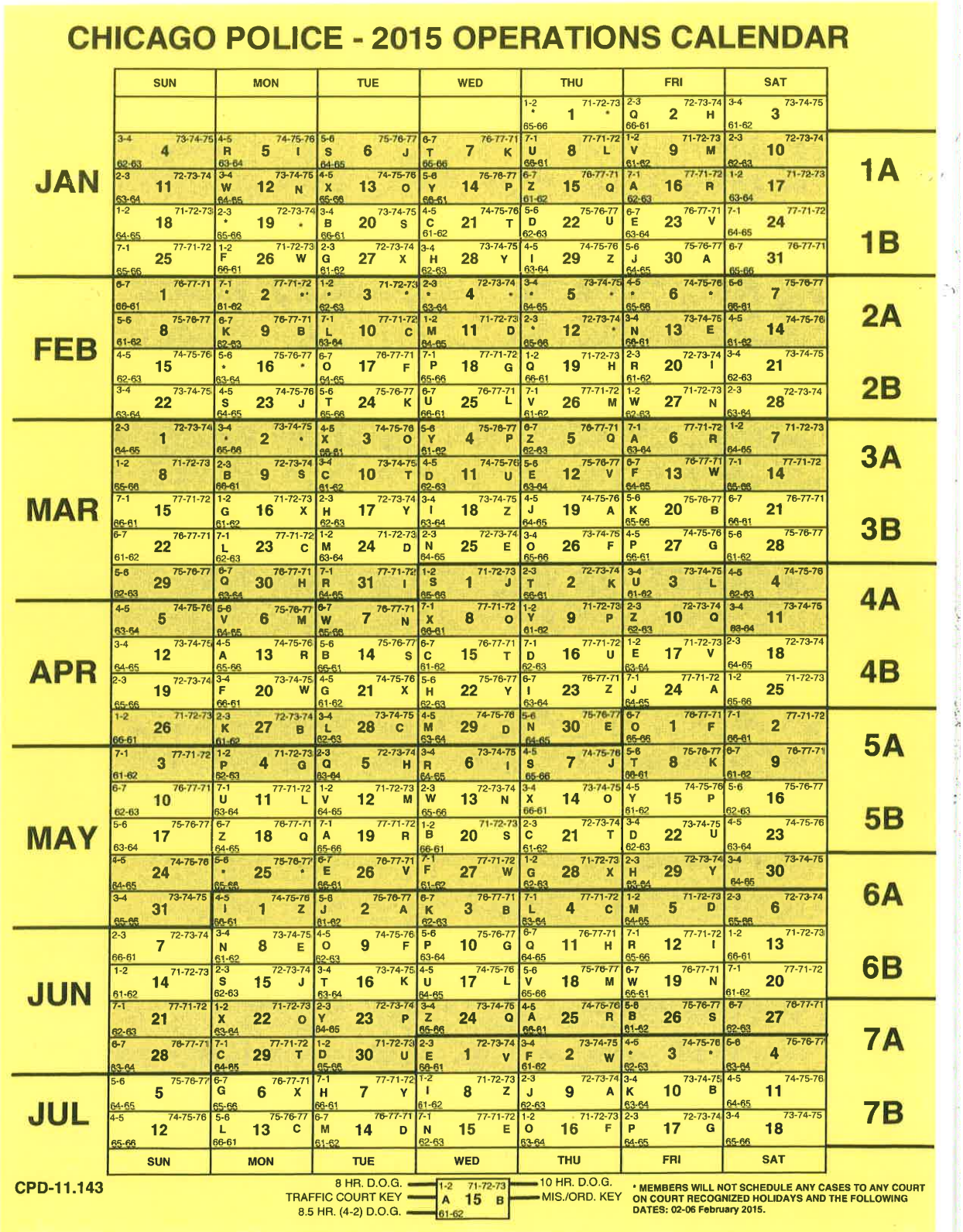
(b) Frac. of Days as field LT



(c) Frac. of Days as Tact LT

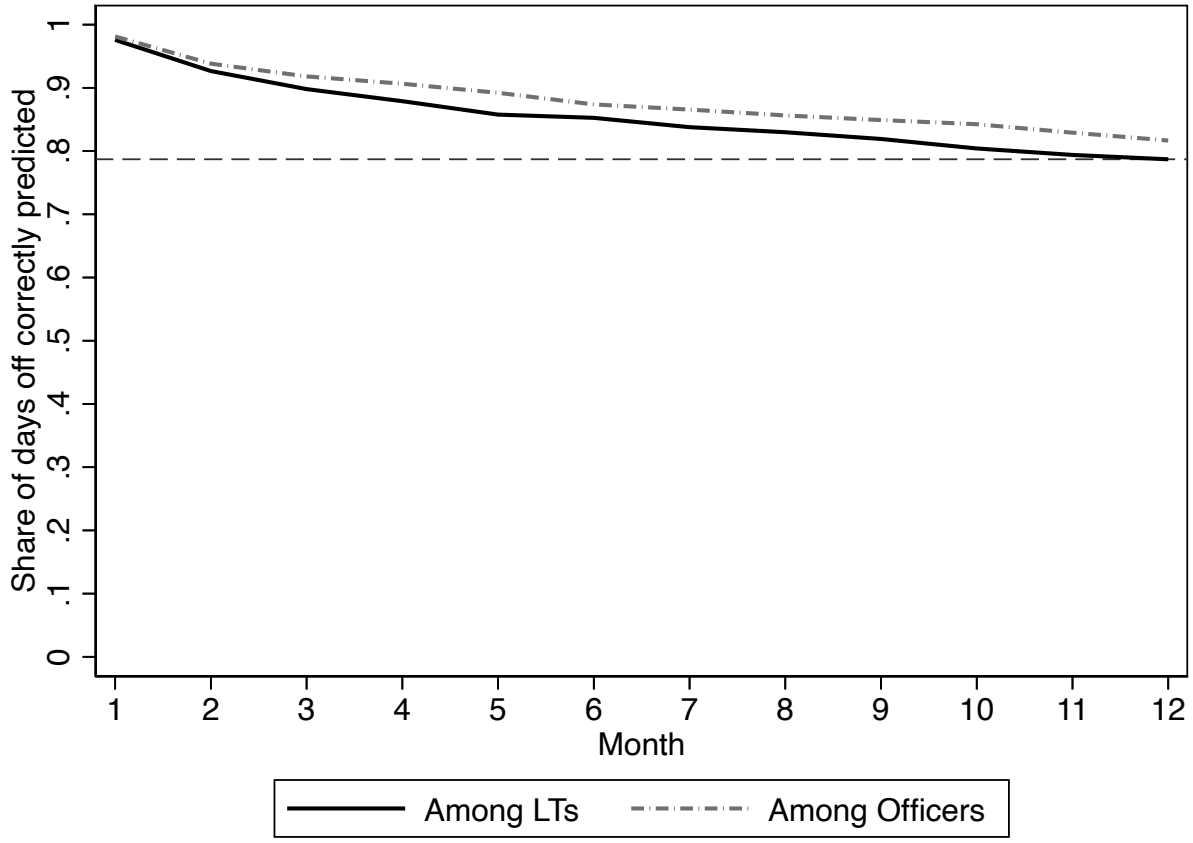
Note: This figure shows how lieutenants who work as watch operations lieutenants split their work days. In the first figure, we show distributions of work days as watch operations lieutenants (WOL) among lieutenants who ever work as WOLs in a month. Specifically, the data are at the lieutenant-month 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



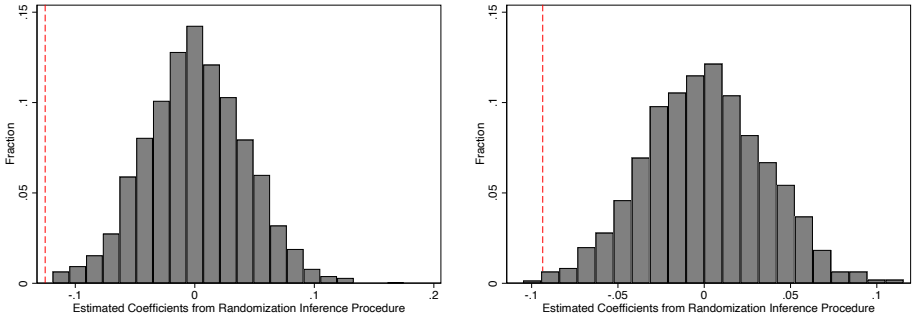
Note: This figure shows the 2015 day-off group calendar obtained from the CPD. The numbers between 61 and 66 in the bottom left corner of each box indicate whether that group has a day off on the given day. For example, on Sunday Jan 4, the 62-63 indicates that day-off groups 62 and 63 are not expected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



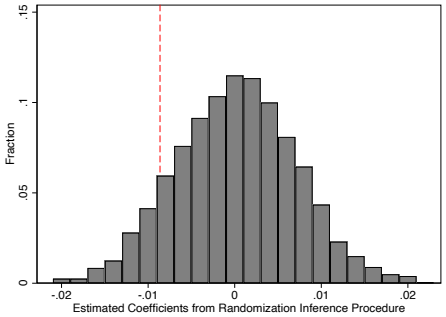
Note: For this figure, we first predict days off for both LTs and officers based only their January attendance records. Then, we merge on officers and LTs actual recorded days off based on attendance records and calculate the fraction of days off that we predict correctly in each month. The solid line 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

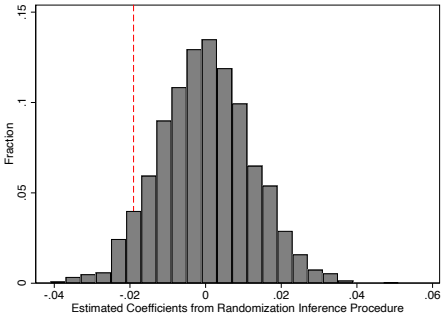


(a) Arrests

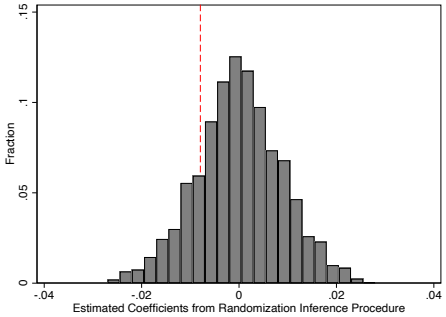
(b) Arrests of Black Persons



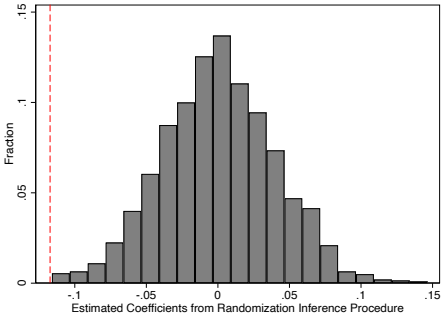
(c) Arrests of White Persons



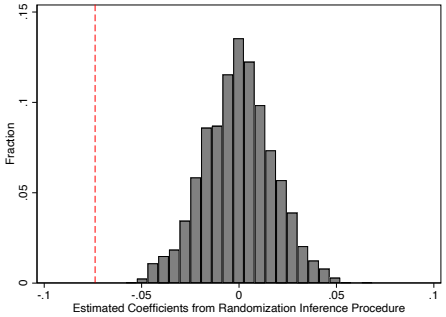
(d) Arrests of Hispanic Persons



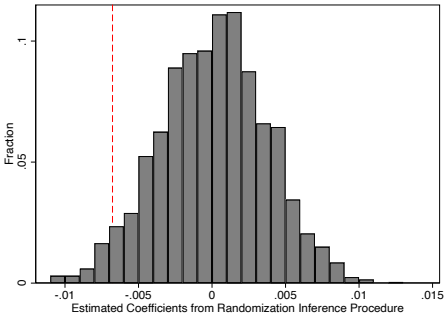
(e) Index Arrests



(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.

Appendix Tables

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	0.451***
	(0.027)
Observations	168,462
Y mean	.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	0.3516	-0.0679	0.0997	-0.0161
	(0.4517)	(0.1412)	(0.2356)	(0.0794)
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	-0.1695	-0.0188	-1.8309	-0.3461
	(0.1523)	(0.0245)	(3.9861)	(0.4674)
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	0.0027	-0.0005	-0.0036	-0.0057
	(0.0073)	(0.0084)	(0.0083)	(0.0092)
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

	(1)	(2)	(3)	(4)
			PANEL A: ARRESTS	
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-0.278***	-0.207***	-0.019	-0.042
	(0.099)	(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	-0.018	-0.260***	-0.164***	-0.015*
	(0.024)	(0.091)	(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: 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	-0.122***	-0.092***	-0.008	-0.018	-0.009	-0.113***	-0.074***	-0.005
	(0.047)	(0.035)	(0.009)	(0.015)	(0.011)	(0.043)	(0.020)	(0.004)
Any Pred. Male LT	-0.035	-0.031	-0.002	-0.002	-0.001	-0.034	-0.011	0.000
	(0.037)	(0.031)	(0.007)	(0.011)	(0.009)	(0.035)	(0.020)	(0.004)
Pred. Num. of LTs	0.031	0.000	0.010**	0.017*	-0.015**	0.046*	0.015	0.003
	(0.028)	(0.023)	(0.005)	(0.009)	(0.006)	(0.026)	(0.015)	(0.003)
Any Pred LT promoted after 2004	-0.008	-0.005	-0.001	-0.001	0.006	-0.014	0.002	-0.009
	(0.049)	(0.037)	(0.010)	(0.015)	(0.010)	(0.048)	(0.029)	(0.009)
Any Pred LT with tenure $>$ 20	-0.056*	-0.028	-0.011*	-0.016	0.008	-0.064**	-0.005	-0.008
	(0.031)	(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	0.023	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	0.016	0.027	0.001	-0.007	0.006	0.010	-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.6: Robustness to Additional Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: ADDITIONAL CALENDAR TIME FIXED EFFECTS				
Any Pred. Black/Hisp. LT	-0.126***	-0.095***	-0.009	-0.019	-0.008	-0.118***	-0.073***	-0.007*
	(0.045)	(0.034)	(0.009)	(0.014)	(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	-0.123***	-0.093***	-0.008	-0.018	-0.008	-0.116***	-0.074***	-0.007*
	(0.043)	(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	-0.121***	-0.090***	-0.009	-0.019	-0.010	-0.111***	-0.076***	-0.006*
	(0.041)	(0.032)	(0.009)	(0.014)	(0.010)	(0.038)	(0.018)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
				PANEL D: CONTROLLING FOR EVERYTHING IN PANELS A-C				
Any Pred. Black/Hisp. LT	-0.119***	-0.089***	-0.009	-0.017	-0.010	-0.109***	-0.074***	-0.006
	(0.040)	(0.032)	(0.009)	(0.014)	(0.010)	(0.037)	(0.018)	(0.004)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168,462	168,462	168,462	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table explores robustness of the results in Table 4 to i) additionally including day-of-the-year 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.7: Controlling for Predicted Officer and Sergeant Demographics, Showing Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.117***	-0.086***	-0.008	-0.019	-0.007	-0.110***	-0.073***	-0.006*
	(0.042)	(0.033)	(0.009)	(0.014)	(0.011)	(0.038)	(0.018)	(0.004)
Officer and Serg. Share Black	-0.576**	-0.649***	-0.030	0.116*	-0.056	-0.520**	0.049	-0.035
	(0.254)	(0.226)	(0.049)	(0.061)	(0.081)	(0.248)	(0.141)	(0.032)
Officer and Serg. Share Hisp.	-0.461*	-0.527***	0.005	0.059	-0.089	-0.372	0.022	-0.022
	(0.240)	(0.202)	(0.052)	(0.084)	(0.058)	(0.228)	(0.130)	(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.8: Restricting to District-watch-Years with Variation in Any Predicted Black/Hispanic LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.144***	-0.121***	-0.006	-0.015	-0.011	-0.133***	-0.078***	-0.008*
	(0.055)	(0.044)	(0.009)	(0.017)	(0.012)	(0.051)	(0.024)	(0.004)
Y mean	3.979	3.037	0.304	0.608	0.551	3.428	1.148	0.076
Observations	42449	42449	42449	42449	42449	42449	42449	42449
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.9: White LTs: Restricting to District-watch-Years with Variation in Any Predicted Black/Hispanic LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. White LT	0.145**	0.128**	0.010	0.005	0.010	0.135**	0.076***	0.016***
	(0.064)	(0.057)	(0.011)	(0.016)	(0.014)	(0.060)	(0.022)	(0.005)
Y mean	3.979	3.037	0.304	0.608	0.551	3.428	1.148	0.076
Observations	42449	42449	42449	42449	42449	42449	42449	42449
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table is analogous to the one in Table A.8 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.10: Alternate Inference

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: BASELINE				
Any Pred. Black/Hisp. LT	-0.125***	-0.094***	-0.009	-0.019	-0.008	-0.117***	-0.074***	-0.007*
	(0.044)	(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	-0.125**	-0.094**	-0.009	-0.019	-0.008	-0.117**	-0.074***	-0.007
	(0.053)	(0.036)	(0.008)	(0.017)	(0.012)	(0.046)	(0.021)	(0.004)
				PANEL C: RANDOMIZATION INFERENCE				
Any Pred. Black/Hisp. LT	-0.125***	-0.094***	-0.009	-0.019	-0.008	-0.117***	-0.074***	-0.007*
Rand. inference se	(0.042)	(0.035)	(0.007)	(0.012)	(0.009)	(0.040)	(0.018)	(0.004)
Rand. inference p-value	[0.0015]	[0.0060]	[0.2150]	[0.1180]	[0.3645]	[0.0025]	[0.0000]	[0.0630]
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
Observations	168462	168462	168462	168462	168462	168462	168462	168462

Note: This Table reports the same estimates as in Table 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.11: Poisson Maximum Likelihood Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.031***	-0.029***	-0.027	-0.029	-0.013	-0.033***	-0.064***	-0.087*
	(0.010)	(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	-0.107	-0.072	-0.009	-0.019	-0.007	-0.100	-0.064	-0.006
p-value	0.003	0.005	0.346	0.199	0.487	0.002	0.000	0.075

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.12: Disaggregating Race of LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: DISAGGREGATING LT RACE				
Any Predicted Black LT	-0.081	-0.061	-0.007	-0.010	-0.015	-0.067	-0.039	-0.005
	(0.083)	(0.069)	(0.012)	(0.018)	(0.017)	(0.078)	(0.029)	(0.005)
Any Predicted Hispanic LT	-0.130***	-0.094**	-0.009	-0.023	-0.004	-0.127***	-0.088***	-0.006
	(0.049)	(0.037)	(0.011)	(0.018)	(0.012)	(0.046)	(0.023)	(0.005)
Y mean	3.529	2.501	0.326	0.670	0.514	3.016	1.015	0.068
P-value for Black LT = Hisp LT	0.613	0.677	0.916	0.587	0.577	0.516	0.164	0.864
				PANEL B: DISAGGREGATING WITH CONTROLS				
Any Predicted Black LT	-0.106	-0.079	-0.010	-0.014	-0.019	-0.088	-0.047*	-0.004
	(0.072)	(0.060)	(0.011)	(0.019)	(0.017)	(0.067)	(0.028)	(0.005)
Any Predicted Hispanic LT	-0.111**	-0.079**	-0.007	-0.021	-0.004	-0.106***	-0.085***	-0.006
	(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
P-value for Black LT = Hisp LT	0.959	0.993	0.833	0.784	0.456	0.810	0.255	0.812
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.13: Share Black/Hisp LT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
				PANEL A: SHARE BLACK/HISP LT				
Share Black/Hisp. LT	-0.176***	-0.135***	-0.017	-0.021	-0.009	-0.167***	-0.099***	-0.011**
	(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	168462	168462	168462	168462	168462	168462	168462	168462
				PANEL B: DUMMIED OUT SHARE BLACK/HISP LT				
$0 <$ Share Black/Hisp. LT ≤ 0.5	-0.082*	-0.059*	0.000	-0.019	-0.006	-0.076*	-0.053***	-0.002
	(0.045)	(0.033)	(0.009)	(0.017)	(0.012)	(0.039)	(0.019)	(0.004)
$0.5 <$ Share Black/Hisp. LT ≤ 1	-0.178***	-0.135***	-0.019*	-0.020	-0.010	-0.167***	-0.098***	-0.011**
	(0.058)	(0.049)	(0.011)	(0.016)	(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.14: Arrests Linked to Court Conviction

	(1)	(2)	(3)	(4)
PANEL A: OVERALL GUILTY ARRESTS TO ALL ARRESTS				
	All			
Difference in Ratio	0.005			
	[0.474]			
Mean Ratio	0.567			
PANEL B: GUILTY ARRESTS RELATIVE TO ALL ARRESTS				
	Index	Non-index	Public Order	Discretionary
Difference in Ratio	0.008*	-0.003	-0.009	-0.001
	[0.054]	[0.603]	[0.114]	[0.396]
Mean Ratio	0.108	0.459	0.155	0.015
PANEL C: GUILTY ARRESTS RELATIVE TO ALL ARRESTS WITHIN CATEGORY				
	Index	Non-index	Public Order	Discretionary
Difference in Ratio	0.001	0.002	0.010	0.044
	[0.935]	[0.742]	[0.527]	[0.441]
Mean Ratio	0.772	0.533	0.524	0.759
Observations	84,231	84,231	84,231	84,231
Baseline FE and controls	Y	Y	Y	Y

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

Table A.15: Other Outcomes

	(1)	(2)	(3)	(4)
	Stops	Use of Force	Complaints	Force Complaints
Any Pred. Black/Hisp. LT	-0.1435	-0.0034	0.0005	-0.0003
	(0.1245)	(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.16: Crime Outcomes

	(1)	(2)	(3)	(4)
	PANEL A: BOTH UNCLEARED AND CLEARED CRIMES			
	All Crime	Index	Non-index	Public Order
Any Pred. Black/Hisp. LT	-0.0894*	-0.0404	-0.0490	-0.0010
	(0.0459)	(0.0291)	(0.0349)	(0.0119)
Observations	168,462	168,462	168,462	168,462
Y mean	11.0980	4.6125	6.4856	1.1146
Baseline FE and controls	Y	Y	Y	Y
	PANEL B: REMOVING CRIMES WITH ARRESTS			
	All Crime	Index	Non-index	Public Order
Any Pred. Black/Hisp. LT	-0.0425	-0.0358	-0.0067	0.0119
	(0.0405)	(0.0281)	(0.0264)	(0.0092)
Observations	168,462	168,462	168,462	168,462
Y mean	8.5925	4.1187	4.4737	0.6650
Baseline FE and controls	Y	Y	Y	Y

Note: This table runs the same regression as in Table 4 for crime outcomes. The first tab reports results using outcomes for crimes that are both cleared and uncleared. The second tab reports results using outcomes that remove crimes that led to arrests.

Table A.17: Controlling for Actual Officer and Sergeant Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.121***	-0.090***	-0.008	-0.020	-0.009	-0.112***	-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: 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.18: Individual-Level Arrests

	(1)	(2)	(3)	(4)
		PANEL A: ARRESTS		
	Arrests	Black Arrests	White Arrests	Hispanic Arrests
Any Pred. Black/Hisp. LT	-.00173**	-.00113**	-.0000367	-.000545*
	(.000726)	(.000525)	(.000144)	(.000308)
Y mean	.0481	.0329	.0048	.00994
FE	Baseline	Baseline	Baseline	Baseline
Any Pred. Black/Hisp. LT	-.00133**	-.000946**	.0000492	-.00044*
	(.000561)	(.000416)	(.000128)	(.000261)
Y mean	.0481	.0329	.0048	.00994
FE	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer
		PANEL B: ARREST TYPE		
	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	.0000877	-.00181***	-.00106***	-.0000644
	(.000163)	(.000684)	(.000331)	(.000059)
Y mean	.00756	.0405	.0138	.000971
FE	Baseline	Baseline	Baseline	Baseline
Any Pred. Black/Hisp. LT	.0000604	-.00139***	-.000984***	-.0000647
	(.000153)	(.00052)	(.000289)	(.0000557)
Y mean	.00756	.0405	.0138	.000971
FE	Baseline + Officer	Baseline + Officer	Baseline + Officer	Baseline + Officer

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

Table A.19: Stops in Various Neighborhoods

	(1)	(2)	(3)	(4)
	Stops	High Black	High White	High Hispanic
Any Pred. Black/Hisp. LT	-0.1435	-0.0850	-0.0511	-0.0257
	(0.1245)	(0.0576)	(0.0402)	(0.0422)
Observations	168,462	168,462	168,462	168,462
Y mean	11.4984	3.6958	2.3005	3.0928
Baseline FE and controls	Y	Y	Y	Y

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

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**	-0.125**	-0.034***	-0.020	-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	-0.061	-0.046	0.004	-0.016	0.005	-0.065	-0.063***	-0.003
	(0.043)	(0.035)	(0.010)	(0.013)	(0.011)	(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
Baseline FE and controls	Y	Y	Y	Y	Y	Y	Y	Y
P-value for Watch LT = Field LT	0.152	0.238	0.014	0.899	0.175	0.212	0.784	0.882

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

Table A.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***	0.112**	0.034***	0.031	0.024*	0.160***	0.088***	0.010*
	(0.066)	(0.051)	(0.011)	(0.020)	(0.014)	(0.059)	(0.027)	(0.005)
Any Pred. White Field LT	0.007	0.011	-0.003	0.002	-0.009	0.017	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	0.057	0.004	0.251	0.054	0.026	0.034	0.075

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

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Arrests	Black Arrests	White Arrests	Hispanic Arrests	Index	Non-index	Public Order	Discretionary
Pred. Black/Hisp. Watch Op. LT	-0.314***	-0.188**	-0.057**	-0.059*	-0.017	-0.297***	-0.159***	0.008
	(0.109)	(0.085)	(0.024)	(0.033)	(0.027)	(0.102)	(0.051)	(0.010)
Pred. Black/Hisp. Field LT	-0.053	-0.044	0.038	-0.046	-0.050*	-0.003	-0.034	0.001
	(0.117)	(0.079)	(0.026)	(0.046)	(0.027)	(0.112)	(0.078)	(0.018)
Pred. Black/Hisp. Tactical LT	-0.007	-0.002	0.002	-0.000	0.059*	-0.066	0.048	0.005
	(0.168)	(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	Y	Y	Y	Y	Y	Y	Y	Y
P-val Watch = Field	0.103	0.202	0.009	0.819	0.378	0.056	0.184	0.728
P-val Watch = Tact	0.140	0.241	0.146	0.333	0.085	0.236	0.079	0.889
P-val Tact = Field	0.861	0.825	0.490	0.605	0.052	0.806	0.633	0.912
P-val Joint Test	0.086	0.241	0.018	0.583	0.139	0.066	0.030	0.899

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: Release without Charge

	(1)	(2)	(3)	(4)	(5)
	All	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	0.003	0.003	-0.000	0.002	0.000
	(0.004)	(0.002)	(0.003)	(0.002)	(0.000)
Y mean	0.053	0.028	0.025	0.006	0.000
Difference in Ratios	0.001	0.007	0.000	0.003	0.002
	[0.222]	[0.141]	[0.825]	[0.142]	[0.446]
Mean Ratio	0.015	0.054	0.008	0.006	0.000
Observations	168,462	168,462	168,462	168,462	168,462
Baseline FE and controls	Y	Y	Y	Y	Y

Note: This table explores how Black and Hispanic LTs affect the number and share of arrests released without charge (RWOC). The first panel looks at the total number of RWOC decisions (overall and for separate arrest categories). In the second panel, we test if there is a difference in the ratio of RWOC over total arrests. Specifically, we estimate coefficients in regressions where the dependent variables are RWOC decisions and arrests. We add the estimated coefficients to their respective dependent variable means and compare the resulting ratio with the ratio of the dependent variable means. The p-values are in brackets. We repeat this exercise for the different arrest categories, examining the ratio of RWOC within an arrest type to the total number of arrests within that type.

Table A.24: Arrests Made by Overtime Officers

	(1)	(2)	(3)	(4)	(5)	(6)
	Overtime Officers	Arrests	Index	Non-index	Public Order	Discretionary
Any Pred. Black/Hisp. LT	-0.049	0.002	0.006***	-0.004	-0.006**	-0.001
	(0.047)	(0.007)	(0.002)	(0.006)	(0.003)	(0.001)
Y mean	1.989	0.104	0.029	0.075	0.024	0.002

Note: In the first column, we show results for whether there are differences in the number of officers who worked overtime under predicted Black and Hispanic LTs. We examine overtime requests made for extension of the tour of duty. In the subsequent columns, we estimate effects on the number of arrests (overall, and then by category) made by the officers who worked overtime. We define overtime work as arrests made within the window of four hours before the start of the overtime and the end of the overtime. The pre-period before the start of the overtime is included because overtime requests are often made to process arrests that are made toward the end of duty.

Table A.25: Arrest Performance and Awards

	(1)	(2)	(3)	(4)	(5)	(6)
	Awards from LTs	Awards from LTs	Awards from LTs	Awards from LTs	Awards from LTs	Awards from LTs
Any Pred. Black/Hisp. LT	0.007	0.012	0.008	0.011	0.013	0.009
	(0.010)	(0.009)	(0.010)	(0.009)	(0.011)	(0.010)
Arrests		0.005***				
		(0.001)				
Arrests × Black/Hisp. LT		-0.001				
		(0.001)				
Index			0.016***			
			(0.002)			
Index × Black/Hisp. LT			-0.001			
			(0.006)			
Non-index				0.003***		
				(0.001)		
Non-index × Black/Hisp. LT				-0.001		
				(0.001)		
Public Order					0.003*	
					(0.002)	
Public Order × Black/Hisp. LT					-0.006*	
					(0.003)	
Discretionary						0.005
						(0.007)
Discretionary × Black/Hisp. LT						-0.027*
						(0.014)
Y mean	0.049	0.049	0.049	0.049	0.049	0.049
Observations	168,462	168,462	168,462	168,462	168,462	168,462

Note: This table presents results on how Black and Hispanic LTs differ in their award nominations for a given level of performance. We use award data from 2013 to 2020, focusing on award nominations for incidents that occurred on a specific day (as opposed to over multiple days) and were specifically given out by LTs. In column (1), we compare the total number of awards given on days with predicted Black and Hispanic LTs to the other days. In the following columns, we regress the number of award nominations on the total number of arrests of certain types, our indicator variable for a predicted Black or Hispanic LT, and the interaction of these two variables, as well as the usual baseline FEs. The coefficient on the arrest category indicates how many more award nominations you get from an additional arrest of the given category. The interaction term indicates how this relationship changes on days with a predicted Black or Hispanic LT.