New funds, new scrutiny for Africa's Great Green Wall p. 666

Do microplastics affect human health? p. 672

Imaging Earth's crust with whale calls p. 731

Science

$15
12 FEBRUARY 2021
sciencemag.org
AAAS

DIVERSITY AND POLICING

Big data reveal threats to minorities policed by white and male officers pp. 677 & 696

RESEARCH ARTICLE

CRIMINAL JUSTICE

The role of officer race and gender in police-civilian interactions in Chicago

Bocar A. Ba[1], Dean Knox[2]*, Jonathan Mummolo[3]*, Roman Rivera[4]

Diversification is a widely proposed policing reform, but its impact is difficult to assess. We used records of millions of daily patrol assignments, determined through fixed rules and preassigned rotations that mitigate self-selection, to compare the average behavior of officers of different demographic profiles working in comparable conditions. Relative to white officers, Black and Hispanic officers make far fewer stops and arrests, and they use force less often, especially against Black civilians. These effects are largest in majority-Black areas of Chicago and stem from reduced focus on enforcing low-level offenses, with greatest impact on Black civilians. Female officers also use less force than males, a result that holds within all racial groups. These results suggest that diversity reforms can improve police treatment of minority communities.

Racial disparities in police-civilian interactions and high-profile incidents of excessive force continue to fuel allegations of abusive and discriminatory policing (*1*, *2*). Central to these critiques are the fact that throughout the history of policing in the United States, many police forces have been nearly all white and male (*3*). In turn, some of the most frequently proposed reforms aimed at reducing inequities and police brutality have centered on hiring more nonwhite (*4*) and female (*5*) officers. One agency that has undergone substantial diversification in recent decades is the Chicago Police Department (CPD), transforming from a mostly white and nearly all male force to one in which half of sworn officers are minorities and over one-fifth are female. This heterogeneity across race and gender lines, combined with newly acquired data on officers' daily patrols and enforcement activities, allows a thorough assessment of the practical consequences of diversity in law enforcement. Although we cannot directly infer the future impact of further diversification, we can examine the Chicago case in depth to provide the most credible microlevel evidence to date on the treatment civilians can expect when encountering officers of varied racial, ethnic, and gender identities.

Theories of social distance and intergroup relations in a range of contexts (*6*–*9*) imply that diversifying police agencies may improve the treatment of minorities (*3*, *8*). Individuals rely on stereotypes when evaluating members of social groups (*10*) and are thought to be less likely to engage in harassment toward in-group members (*11*). However, research on organizational culture and bureaucratic politics suggests that officers of different social backgrounds may ultimately behave similarly because of self-selection into service and socialization during training and on the job (*3*, *12*–*16*). To succeed and advance, women and minorities may also face pressure to adopt conventional enforcement practices (*3*, *17*, *18*).

Rigorous evaluation of the effects of police diversity has been stymied by a lack of sufficiently fine-grained data on officer deployment and behavior that makes it difficult or impossible to ensure that officers being compared are facing common circumstances while on duty. Studies typically rely on coarse geographic units, like agency- or precinct-level data (*19*–*21*), which forced previous scholars to invoke the strong assumption that, for example, "white and nonwhite officers are randomly assigned to neighborhoods" (*20*, p. 389). Furthermore, most policing data sets contain records of enforcement events only [e.g., logs of stops or arrests (*22*–*24*)]; events in which officers choose to take no action are unobserved, potentially distorting inferences. Other studies that make valid comparisons are often limited in scope to particular activities, like ticketing during traffic accident investigations (*25*). And although some prior work has leveraged the timing of diversity reforms to estimate agency-level effects (*26*–*29*), those aggregated approaches are by design unable to examine details of police-civilian interactions. Findings with regard to racial diversity in particular have been decidedly mixed: In an exhaustive review of the empirical literature, one prominent legal scholar concluded, "[t]he fairest summary of the evidence is probably that we simply do not know" (*30*).

To assess the impact of diversity in law enforcement, we draw on newly collected data, assembled through years of open-records requests, that allow us to overcome long-standing limitations. These include officer demographics, language skills, daily shift assignments, and career progression. We link these files to time-stamped, geolocated records of the same officers' decisions to stop, arrest, and use force against civilians. After aggressively pruning data to maximize analytic validity, we compile a panel of 2.9 million officer shifts and 1.6 million enforcement events by nearly 7000 officers covering the years 2012 through 2015. Most notably, we leverage fine-grained information on daily patrol assignments, which vary exogenously on the basis of fixed rules and preassigned rotations, to examine how officers of different groups behave when faced with comparable circumstances and civilian behaviors.

The deployment effects that we estimate are a critical first step in the systematic evaluation of widely proposed personnel reforms, which have historically focused on increasing racial and gender diversity among officers. If officers of different demographic profiles do not behave differently when faced with the same conditions, there is little hope that diversifying police agencies will yield tangible differences in the treatment of marginalized civilians. Indeed, we demonstrate that deploying officers of different demographic profiles to comparable environments does produce large differences in how police treat civilians. However, we caution that these deployment effects do not directly generalize to future effects of hiring reforms, for several reasons. Chief among these are that (i) the nature of police-civilian interactions is changing rapidly; (ii) racial, ethnic, and gender differences in current officers' behavior may not map perfectly to those of future cohorts; (iii) deployment patterns will necessarily change as more officers are hired from marginalized groups; and (iv) diversification reforms may exert additional, potentially powerful second-order effects, e.g., through agency culture.

Chicago as a case study

Our focus on one city provides unusually detailed data at the expense of geographic scope. Chicago is a large and racially diverse metropolis, with roughly half of residents identifying as nonwhite. Chicago is also heavily segregated, has a history of racial tensions between residents and police, and has come under recent scrutiny for controversial aggressive policing tactics such as "stop and frisk" (*31*). The agency received national attention for the 2014 killing of 17-year-old Laquan McDonald, an attempted cover-up, and ensuing social unrest (*32*). The CPD was condemned for its "code of silence" (*33*), and then-superintendent Garry McCarthy received widespread criticism for "encouraging

[1]Department of Economics, University of California Irvine, Irvine, CA, USA. [2]Operations, Information and Decisions Department, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Politics, School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [4]Department of Economics, Columbia University, New York, NY, USA.
*Corresponding author. Email: dcknox@upenn.edu (D.K.); jmummolo@princeton.edu (J.M.)

the kind of aggressive cop culture under which McDonald's shooting took place" (*34*).

As (*26*) recounts, before a series of lawsuits in the 1970s, the CPD was slightly less than 20% Black, in a city that was one-third Black in 1970. The Afro-American Patrolmen's League (AAPL) filed a discrimination suit "on hiring, promotion, assignment, and discipline" (*35*), with the Department of Justice (DOJ) soon joining it (*26*). In the early 1970s, Black hiring shares were around 10%. In 1974, hiring quotas were imposed, and Black hiring shares increased to 40% by 1975 (*26*). These reforms had broader effects on CPD's composition of the department; women made up a larger proportion of Black recruits, with white women lagging in hiring and promotion at first (*36*). As of December 2016, roughly 22% of officers identified as Black, 23% as Hispanic, and 3% Asian; and 22% are female, a stark change from its 99% male workforce in 1970. (Text S1.1 discusses racial and ethnic classification of CPD officers; text S2.1 and fig. S1 provide additional details about the CPD's evolution.)

From one standpoint, it may be difficult to extrapolate from Chicago to settings lacking these racial tensions and history of diversification. But in other ways, it is these very conditions that make Chicago an important test case: Among major departments nationwide, it is arguably one in which reform has historically been sorely needed. A single case study cannot be the final word in an important debate. But Chicago offers an invaluable opportunity to study diversity in policing using unusually fine-grained data, in a setting where concerns over racial inequity are pronounced.

Data

We submitted a series of open-records requests and appeals to the CPD, the city's Department of Human Resources, and the Illinois Office of the Attorney General over 3 years, seeking data on officer demographics and behavior. The resulting records include the name, race, gender, language skill, unit assignments, and appointment date of each officer (*37*, *38*). We also obtained officers' stops, arrests, and uses of force, which we merged with daily patrol assignments and U.S. Census data, per text S1.2 and S1.3. Table 1 reports aggregate counts. Owing to sparse data on other groups, our analysis is limited to Black, Hispanic, and white officers (97% of officers in the sample). Stops and arrests are recorded in officer-shift data once per officer contributing to enforcement. In stop records, one is listed as "first" officer, suggesting a leading role, although arrest records contain no such labeling. (Text S3.9 and fig. S2 conduct additional analyses of first officers only, yielding a highly similar pattern of results.)

Figure 1 depicts a small slice of the data's temporal and geographic coverage: a 3-month window in CPD's Wentworth District (District 2), a highly segregated 7.5-square-mile territory on Chicago's South Side that is 95% Black and consistently ranks among the city's most violent districts in per-capita crime. The district spans 15 patrol areas, shaded according to racial composition. Points indicate geolocated stops, arrests, and uses of force during this period. The figure also offers a detailed portrait of four anonymized CPD officers working in District 2 in this time. For example, "Officer A" is female, Black, does not speak Spanish, and joined the CPD in 1994; "Officer C" is a white male who joined the CPD in 2006 and does not speak Spanish. The figure shows officers' specific patrol slots and each officer's behavior while on assignment.

Identifying racial, ethnic, and gender disparities in policing

Although the CPD has diversified over time, officer groups face substantially different working conditions. Figure 2 displays the average characteristics of districts—22 geographic regions delineated by the CPD—to which officer groups are assigned. Differences associated with officer race and ethnicity are most stark. In general, Black officers work in districts with 47% higher per-capita violent crime and large co-racial populations—on average 68% co-racial, far higher than the average 26 to 30% co-racial and co-ethnic districts where white and Hispanic officers serve. However, white officers are generally overrepresented relative to the resident population; 20% of the 95%-Black Wentworth District officers are white; in Austin (District 15), where residents are 93% Black, officers are 55% white. (Text S2.2 discusses district organization. Figures S3 to S5 present district-level data; assigned officer demographics somewhat track those of district residents, but officers are disproportionately white.) Even within districts, text S2.3 and figs. S6 and S7 demonstrate that marginalized groups are tasked with patrolling different beats, compared to white or male colleagues. (All p-values < 0.001.)

These patterns underscore a central difficulty in evaluating how officer behavior varies across demographic groups. Namely, white officers work in different environments from minority officers, on average. Men and women also work during different hours of the day (text S2.3 and figs. S9 and S10). This means that after aggregating to large geographic units and time periods, observed behavioral differences may simply reflect differing patrol environments, rather than differences in policing approaches.

To make valid comparisons, we assemble a panel dataset in which rows represent officer-shifts—roughly 8-hour patrol periods—and characterize officers' actions and their context. (Text S1.2 describes these datasets; text S1.4 and S1.5 elaborate on preprocessing.) In each of these 2.9 million patrol assignments, we measure officers' stops, arrests, and uses of force, whether they engaged in any of these activities or not. We compared officers of different demographic profiles working in the same specific combination of month and year (e.g., January 2012), day of week, shift time, and assigned "beat" (a patrol task, typically corresponding to small geographic areas less than one square mile; see text S1.6 for a detailed discussion of beat assignments)—a narrow slice of time and space that we abbreviate "MDSBs" (month, day of week, shift, beat). CPD also assigns officers to "day-off groups," which determine who works on rotating dates according to a scheme set late in each calendar year for the following year, representing a large exogenous source of variation in the officers that are available to serve in a particular patrol assignment on any given date. This procedure greatly mitigates threats from self-selection (e.g., officers choosing to take

Table 1. Summary of data on officer behavior (counts), 2012-2015. Summary statistics after pruning officers, shifts, and event records aggressively to ensure common circumstances in our behavioral analysis.

	Black officers	Hispanic officers	White officers	Female officers	Male officers
Stops	253,576	356,493	729,000	264,526	1,074,543
Arrests	47,396	65,581	132,272	43,625	201,624
Uses of force	1,355	2,081	4,513	1,125	6,824
Shifts	829,818	689,091	1,413,771	740,015	2,192,665
Officers	1,834	1,674	3,439	1,785	5,162

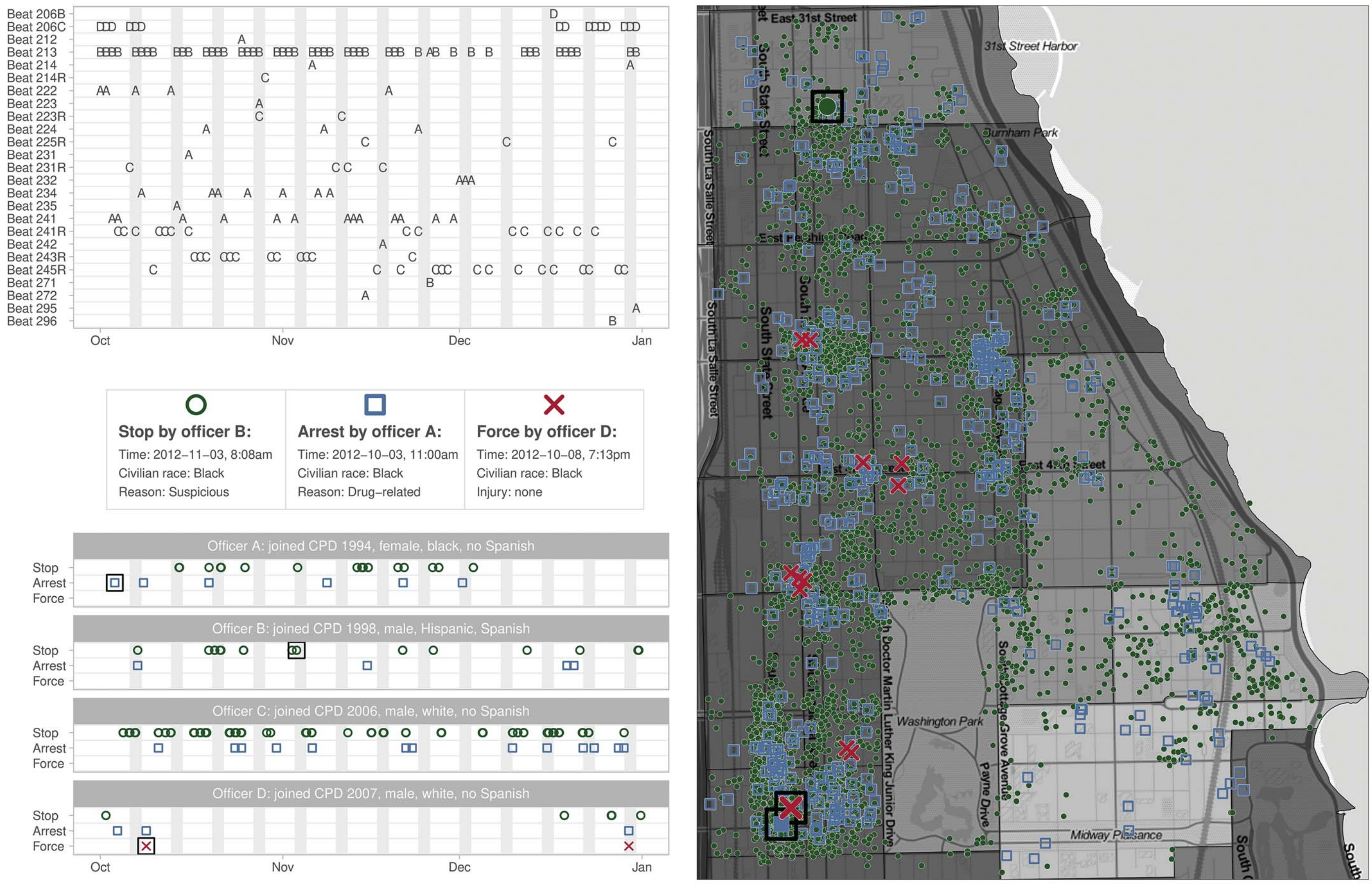

Fig. 1. Detailed view of the data. The right panel maps police activity in a single CPD district (Wentworth, District 2), with green circles, blue squares, and red crosses respectively indicating the locations of stops, arrests, and uses of force. Polygons represent geographic beats and are shaded by their proportion of minority residents. Lower left panels chart the behavior of four anonymized officers over a 3-month period, with panel headers indicating the year officers joined CPD, gender, ethnicity or race, and language ability. Boxed incidents are described further in the left middle panel, which reports civilian and incident specifics. Finally, the top left panel indicates how the four selected officers are assigned to patrol beats over dates and times, with vertical gray bars indicating weekends.

days off when crime spikes), helping to ensure that the officers we compare are facing common circumstances, on average. This design also ensures that comparisons of officer activity have common denominators, as those working in comparable places and times have the same opportunity to take enforcement action. (See text S1.6 and figs. S11 and S12 for further details on CPD shift assignment procedures and related analytic strategy. Text S3.6 and fig. S13 demonstrate how common alternative approaches in prior work can mislead analysts about the magnitude or even the sign of these effects. Text S3.3 examines clock-in/out times by officer race and finds no meaningful differences in shift duration.)

Because our analytic approach relies on comparisons between officers deployed to the same MDSBs, our inferences are limited to MDSBs in which cross-group comparisons are feasible, e.g., in which both Black and white officers are both deployed. (See texts S3.1 and S3.2 for discussions of the data-generating process and statistical estimand, as well as how this analytic strategy circumvents the threats to inference posed by unobserved differences in patrol environments.) Thus, our estimates do not necessarily generalize to every officer, time, or location in Chicago. (Text S3.4 and fig. S8 provide details on the roughly 40% of patrol assignments where cross-group, within-MDSB comparisons are unavailable—typically smaller patrol tasks with fewer assigned officers—because assigned officers are all from the same demographic group. Feasibility does not meaningfully vary with resident racial composition, and nearly every officer rotates through patrol tasks with feasible comparisons.) Officers from different demographic groups also differ in unobserved ways. We therefore estimate the average difference in officer behavior resulting from deploying an officer of one demographic profile—and all the associated traits of that demographic label—relative to another, holding environmental conditions constant. Our results do not reflect the hypothetical effect of changing an officer's race or gender while holding their other traits fixed. Rather, they reflect average differences in treatment that civilians can expect when police commanders assign officers of one demographic group to their temporal and geographic vicinity, compared to another officer group, holding circumstances equal.

We present differences estimated using ordinary least squares with MDSB fixed effects, though our results are robust to several other estimators, including the addition of flexible controls for experience (see text S3 for estimation details and additional results). All statistical inferences are based on officer-level block bootstrap confidence intervals (CIs) that are robust to unobserved officer-specific peculiarities.

Results

Figure 3 displays average differences in the number of stops, arrests, and uses of force by Black and Hispanic officers (relative to white officers) and female officers (relative to male) working in the same MDSBs. Turning first to Black officers, Fig. 3 shows that when faced

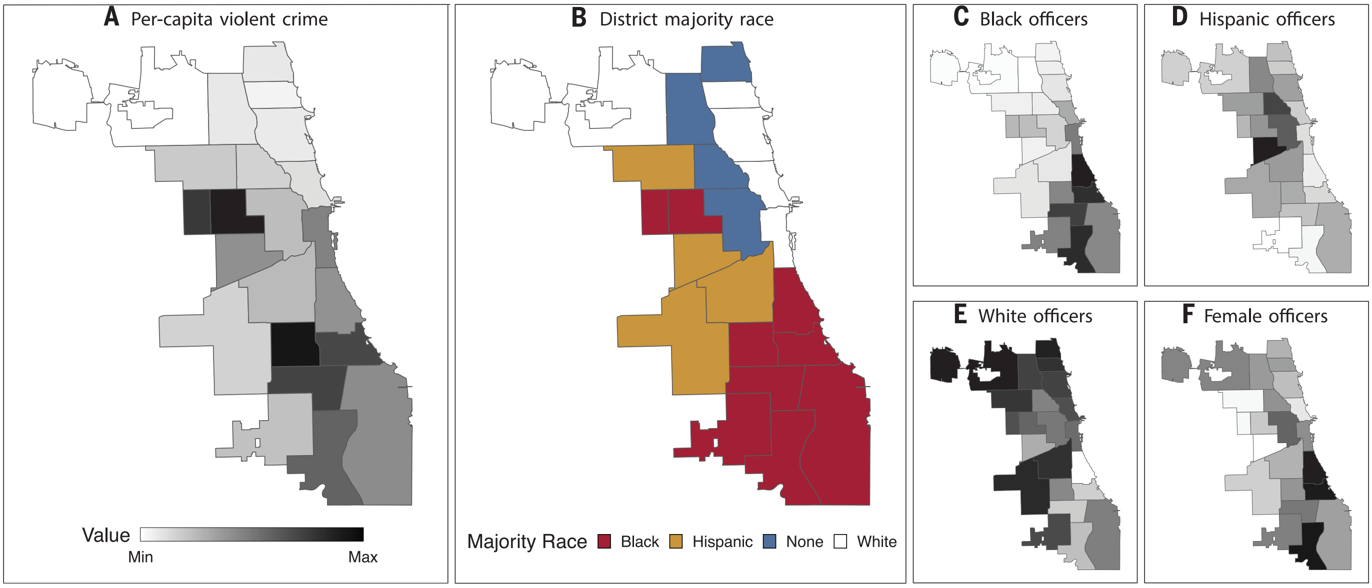

Fig. 2. Unit assignments of various officer groups. (**A** and **B**) Panels depict background information on CPD districts. (**C** to **F**) Panels show, for each district, the proportion of assigned officers belonging to a particular demographic group. Majority race of district residents is based on 2010 decennial Census data; all other plots use 2013–2016 CPD records.

with comparable working conditions over the course of 100 shifts, this group makes 15.16 fewer stops and 1.93 fewer arrests, and it uses force 0.10 fewer times than white counterparts on average—that is, compared to white officers given the same patrol assignment, in the same month, on the same day of the week, and at the same shift time (all $p_{\mathrm{adj}} < 0.001$ after Benjamini-Hochberg multiple-testing correction for all cross-group comparisons and all categories and subcategories of enforcement). These gaps are large, representing 29, 21, and 32% of the average stop, arrest, and use-of-force volume for white officers citywide (see tables S1 to S3 for average enforcement activity by officer group. See tables S4 to S8 for full numeric results, including an additional analysis of Spanish language ability).

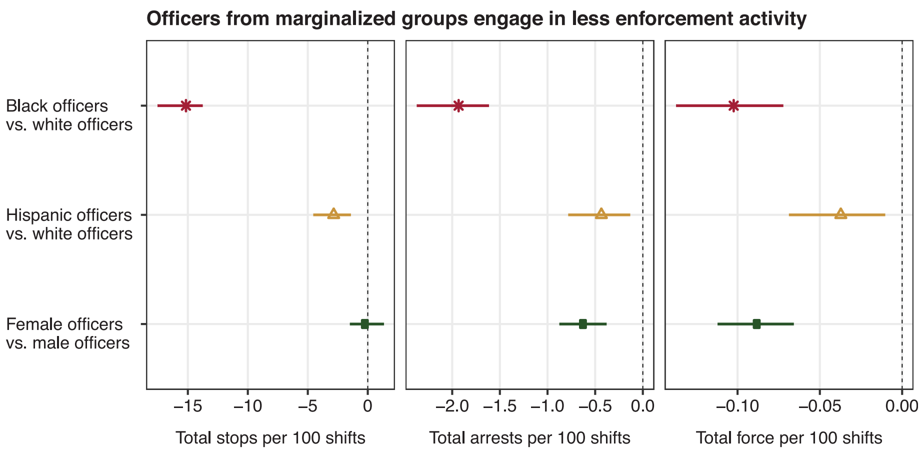

Fig. 3. Effects of deploying officers from marginalized groups on total enforcement activity. Average within-MDSB differences in rates of stops, arrests, and uses of force across officer groups. See tables S4 to S8 for numeric results.

Figure 4 shows that these disparities are not uniform across situations but are driven by a reduced focus on Black civilians. For example, deploying Black officers instead of white yields 12.55 fewer stops of Black civilians per 100 shifts, a reduction equal to 39% of typical white-officer volume. By contrast, Black officers make only 1.31 fewer stops of white civilians per 100 shifts than their white counterparts (reduction equal to 17% of typical white-officer volume; all adjusted p-values < 0.001). The large differences in these scaled effects (39% versus 17%) suggest that they are not explained by the fact that police engage Black civilians in Chicago more often in general. Put another way, in 100 typical white-officer shifts, Black-civilian stops (32.45) are far more frequent than white-civilian stops (7.53), occurring with a baseline ratio of 4.31. By contrast, when deploying Black officers in lieu of white officers, the ratio of the reductions in stops of Black civilians (−12.55) to white civilians (−1.31) is twice as large: 9.60 (95% CI [8.00, 12.98]). Similarly, the ratio of the reductions on Black-civilian arrests (−1.46) to white-civilian arrests (−0.18) is 7.99 (95% CI [5.60, 15.01]). This is significantly larger than the ratio of typical white-officer enforcement volumes (5.90 for Black civilians, 1.17 for white civilians, ratio of 5.03). Black officers also deploy force against Black civilians 0.08 fewer times per 100 shifts than their white counterparts, and they use force that results in injury 0.03 fewer times per 100 shifts (reductions equal to 38 and 39% of typical white-officer volume, respectively; all adjusted p-values < 0.001). Reduced use of force against Black civilians accounts for 83% of the overall force disparity between white and Black officers.

Compared to white officers working in comparable places and times, Black officers also show reduced focus on enforcement activities that are more discretionary in nature. For example, Black officers make 5.72 fewer stops per 100 shifts for "suspicious behavior" (a reduction equal to 31% of average white-officer volume). The reduction resulting from deploying Black officers on drug arrests (−0.31 per 100 shifts) is also estimated to be larger than the effect on violent arrests (−0.23 per 100 shifts). Though the raw effects on drug and violent crime arrest counts are not statistically distinguishable from one another, comparing these effects to typical baseline enforcement

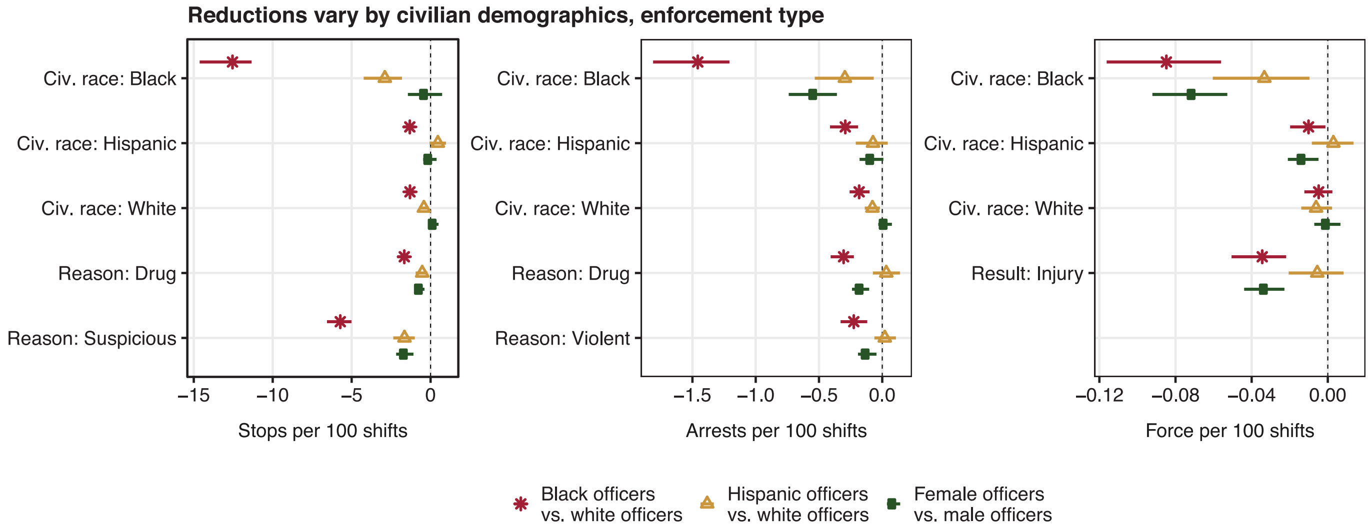

Fig. 4. Effects of deploying officers from marginalized groups, by enforcement subcategory. Average within-MDSB differences in rates of stops, arrests, and uses of force across officer groups, for selected subcategories of each enforcement type. See fig. S14 and tables S4 to S8 for complete subcategory results.

volume puts them in context. The ratio of these reductions, 1.36 (95% CI [0.87, 2.71])—i.e., slightly larger reductions in drug arrests versus violent-crime arrests—is larger than the baseline ratio of drug arrests (1.13 per 100 shifts) to violent arrests (2.16 per 100 shifts) typically made by white officers (ratio of 0.52). (See text S3.5, tables S4 to S8, and fig. S14 for detailed results, including additional enforcement subcategories. Text S3.8 and figs. S15 to S18 show that results obtained with a wide range of alternative estimators are almost identical.)

These patterns are largely in line with the hopes of proponents of racial diversification, who seek to reduce abusive policing and mass incarceration, especially in Black communities.

Like their Black colleagues, Hispanic officers facing the same working conditions conduct fewer stops, make fewer arrests, and use force less than white officers, though the gaps are more modest. Notably, disparities are primarily driven by less engagement with Black civilians; Hispanic officers exhibit nearly the same average volume of enforcement activity against Hispanic civilians as do white officers. Hispanic officers make 2.84 fewer stops per 100 shifts (a reduction equal to 6% of average white-officer volume, $p_{adj} = 0.001$); 0.44 fewer arrests per 100 shifts ($p_{adj} = 0.012$, 5%); and 0.04 fewer uses of force per 100 shifts ($p_{adj} = 0.021$, 12%). We caution that the descriptor "Hispanic" encompasses a range of cultures and national origins that our data do not allow us to parse and that may correspond to important heterogeneity in behavior. (For example, in tables S7 and S8, we show suggestive evidence for differences between Hispanic officers who can and cannot speak Spanish.) More fine-grained data on officers of Hispanic identity are needed to explore this finding.

We also find differences in female officers' behavior relative to male officers, though these are generally smaller in magnitude. Female officers make 0.61 fewer total arrests per 100 shifts (a reduction equal to 7% of average male officer arrests) and 0.54 fewer arrests of Black civilians per 100 shifts (reduction equal to 9% of average male volume, both $p_{adj} < 0.001$). Indeed, about 88% of this disparity in arrest rate is due to reduced arrests of Black civilians. We also find that female officers use force 0.09 fewer times overall (a reduction equal to 28% of average male volume) and 0.07 fewer times per 100 shifts against Black civilians (reduction equal to 31% of average male volume, both $p_{adj} < 0.001$), with the latter accounting for 81% of overall force reduction. (Figure S19 shows that within each racial and ethnic group, female officers use significantly less force than male counterparts.)

Figure 5 displays core results estimated separately in districts where different racial and ethnic resident groups represent majorities. The gap in activity between white and Black officers is most pronounced in majority-Black areas of the city—further evidence that reductions in stops, arrests, and uses of force by Black officers are driven by a reduced focus on Black civilians. Figure S20 illustrates how enforcement differences in these areas are particularly pronounced at night. We see much less heterogeneity across neighborhoods when comparing Hispanic and white officers. (For additional results on gender heterogeneity, see text S3.7 and fig. S19.)

Discussion

Violent and sometimes fatal encounters between white police officers and unarmed racial minorities continue to prompt widespread calls for law enforcement reforms. Protests against police brutality and racial bias remain ubiquitous, including recent unrest in the location of this study, Chicago. Prominent among the many proposed reforms is increasing the level of racial and gender diversity of police agencies. To evaluate the impact of this approach on police-civilian interactions, we leverage unusually rich data on police personnel and activity in Chicago, a jurisdiction that has already instituted diversity reforms.

We first show that minority officers receive vastly different patrol assignments. Without accounting for this disparity, there is no way to meaningfully characterize the differences in behavior across officer groups. In supplemental analyses (text S3.6 and fig. S13) we replicate our core analysis while iteratively imposing data restrictions common in previous analyses to show that common data constraints can lead to severely mistaken inferences, sometimes reversing substantive conclusions entirely. These disparities between analytic approaches suggest one explanation for the at-present mixed conclusions in studies on officer diversity: Data scarcity has imposed severe analytic constraints that can produce divergent, mistaken conclusions.

We account for these differences in working conditions by analyzing officers working in comparable places and times. We show that Black officers are less likely to stop, arrest, and use force against civilians, especially Black civilians, relative to white officers. These disparities are driven by reduced discretionary stops and arrests for petty crimes, including drug offenses, which have long been thought to fuel mass incarceration (*1*). By contrast, Black officers' violent crime enforcement is

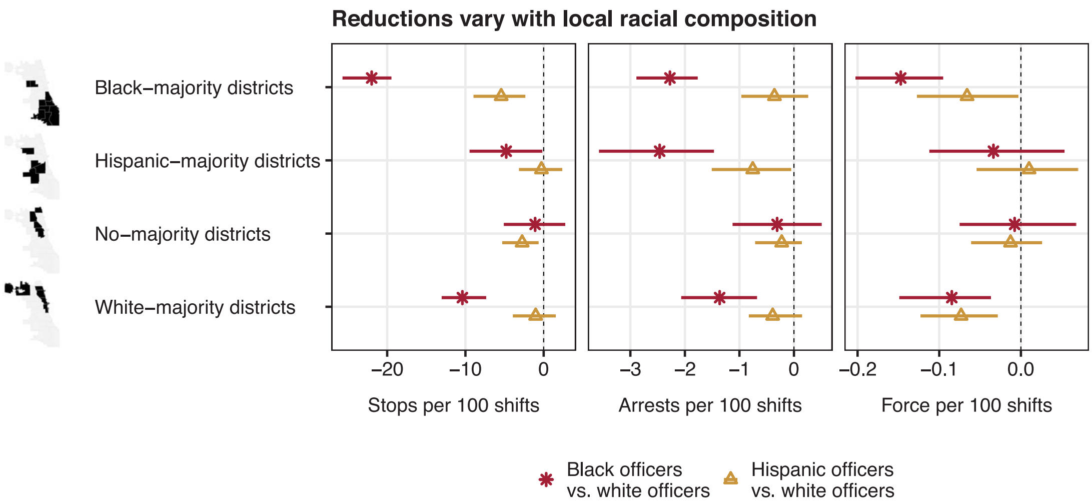

Fig. 5. Effects of deploying officers from marginalized groups, by local racial composition. Average within-MDSB differences in stops, arrests, and uses of force across officer groups, disaggregating CPD districts by majority-resident demographics.

only slightly lower than white officers'. Hispanic officers display lower levels of enforcement activity than whites overall, but their behavior toward Hispanic civilians is broadly comparable to that of white officers, a pattern that deserves further investigation with more fine-grained data on this ethnic group. We also find substantial differences in the behavior of female officers—both relative to male officers generally and within racial and ethnic groups—with the most substantial differences pertaining to use of force. The vast majority of gendered reductions stem from a reduced focus on arresting and using force against Black civilians.

Our results also reveal patterns requiring further study, especially with regard to causal mechanisms. One explanation for these disparities centers on racial bias, i.e., white officers are more likely than Black officers to harass Black civilians. Technically, it is also possible that Black officers respond more leniently when observing crimes in progress (*25*). Though we cannot fully disentangle these observationally equivalent explanations, our data show that these enforcement disparities are predominantly focused on relatively minor crimes, not violent offenses, suggesting little trade-off in terms of public safety. Arbitrating between these competing mechanisms will require objective information on civilian behavior.

Nevertheless, these results help evaluate the promise of proposed personnel reforms by showing what average behavior can be expected when deploying officers of a given demographic profile, relative to their counterparts, holding environmental factors fixed. If we were unable to discern disparities in behavior across these officer groups, diversity reforms would be unlikely to meaningfully alter the volume and character of policing. In fact, not only do we observe differences in enforcement patterns, we also find that these gaps remain nearly identical when adjusting for officer experience (see figs. S15 to S17), an important consideration when extrapolating from retrospective assessments to the future hiring of inexperienced officers. Despite these effects, and even in this highly diverse department, Black and Hispanic civilians in Chicago are engaged by police at rates disproportionate to their shares of the population (though such disparities can arise from a variety of sources, including deployment patterns, civilian behavior, or officer bias). Although our results show that diversity in law enforcement can narrow these gaps, it cannot, on its own, fully address the substantial racial disparities that characterize the American carceral system.

Our analysis uses data from a single city, allowing for an unusually detailed analysis at some cost in generalizability. At present, a patchwork of nonstandard record-keeping and disclosure practices across roughly 18,000 U.S. police agencies (*39*) has severely impeded broader policy evaluations. Our approach, patrol-assignment analyses, offers a useful and widely applicable template for other scholars to follow when testing whether our findings hold in other places and times. But these efforts will require collection of similar data elsewhere, likely necessitating open-records requests, litigation, or data-transparency reforms to compel the release of patrol records that have rarely been shared freely. Acquiring data in these ways can also help mitigate selection bias that can result from forming research partnerships with police agencies, an approach that may skew the literature by focusing on cooperative jurisdictions.

Taken together, these results strongly suggest that diversification can reshape police-civilian encounters. But extrapolation to future hiring hinges on whether recruits come from a comparable pool of potential employees and are deployed in comparable ways. Policing is evolving rapidly, and a complete understanding of the efficacy of reforms requires continued, in-depth research. As officers from marginalized communities increasingly join police forces, their presence will necessarily lead to shifts in deployment and department norms. In turn, shifting deployment patterns may reshuffle officers with particular dispositions to different locations. This could produce different results if, for example, the white officers who are most violent toward Black civilians are then removed from Black neighborhoods, which could shrink the gap in force rates relative to Black officers. If so, the cost-benefit calculus of diversification would be further complicated. The framework that we provide in this study provides a template for future scholars to reevaluate these effects as necessary.

The effects of diversification are likely neither simple nor monolithic. Officers are multidimensional, and crafting effective personnel reforms will likely require thinking beyond the coarse demographic categories typically used in diversity initiatives and consideration of how multiple attributes relate police to the civilians they serve.

REFERENCES AND NOTES

1. M. Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (The New Press, 2010).
2. A. Lerman, V. Weaver, *Arresting Citizenship: The Democratic Consequences of American Crime Control* (Univ. of Chicago Press, 2014).
3. J. Forman Jr., *Locking Up Our Own: Crime and Punishment in Black America* (Farrar, Straus and Giroux, 2017).
4. U.S. Department of Justice, "Advancing Diversity in Law Enforcement" (2016).
5. A. Fantz, C. Tolan, "Want to reform the police? Hire more women," CNN, 2020.
6. Z. Qian, *Race Soc.* **5**, 33–47 (2002).
7. T. A. Laveist, A. Nuru-Jeter, *J. Health Soc. Behav.* **43**, 296–306 (2002).
8. E. M. Kearns, *Justice Q.* **34**, 1213–1245 (2017).
9. H. Tajfel, J. Turner, in *The Social Psychology of Intergroup Relations* W. G. Austin, S. Worchel, Eds. (Wadsworth, 1979), pp. 56–65.
10. C. Neil Macrae, G. V. Bodenhausen, *Annu. Rev. Psychol.* **51**, 93–120 (2000).
11. S. T. Fiske, *Curr. Directions Psychol. Sci.* **11**, 123–128 (2002).
12. J. D. Clinton, D. E. Lewis, *Polit. Anal.* **16**, 3–20 (2008).
13. D. E. Lewis, *Int. Public Manage. J.* **13**, 46–55 (2010).
14. E. Linos, N. Riesch, *Public Admin. Rev.* **80**, 92–103 (2020).

15. J. Q. Wilson, *Bureaucracy: What Government Agencies Do and Why They Do It*, J. Q. Wilson, Ed. (Basic Books 1989).
16. K. J. Meier, *Am. Polit. Sci. Rev.* **69**, 526–542 (1975).
17. V. Ray, *Am. Sociol. Rev.* **84**, 26–53 (2019).
18. S. H. Alim, G. Smitherman, *Articulate While Black: Barack Obama, Language, and Race in the U.S.* (Oxford Univ. Press, 2012).
19. J. Legewie, J. Fagan, "Group threat, police officer diversity and the deadly use of police force," *Columbia Public Law Research Paper* (2016).
20. J. J. Donohue III, S. D. Levitt, *J. Law Econ.* **44**, 367–394 (2001).
21. B. Smith, *Policy Stud.* **31**, 147–162 (2003).
22. A. Fagan, A. Braga, R. K. Brunson, A. Pattavina, *Fordham Urban Law J.* **43**, 539–614 (2016).
23. P. B. Hoffman, E. R. Hickey, *J. Crim. Justice* **33**, 145–151 (2005).
24. E. K. Weisburst, *AEA Papers and Proceedings* **109**, 152–156 (2019).
25. J. West, "Racial Bias in Police Investigations," Working Paper, 2018; https://people.ucsc.edu/~jwest1/articles/West_RacialBiasPolice.pdf.
26. J. McCrary, *Am. Econ. Rev.* **97**, 318–353 (2007).
27. A. R. Miller, C. Segal, *Rev. Econ. Stud.* **86**, 2220–2247 (2019).
28. A. Harvey, T. Mattia, "Reducing Racial Disparities in Crime Victimization," Working Paper, 2019; https://wp.nyu.edu/annaharvey/wp-content/uploads/sites/6417/2020/07/Victimization_Revised.pdf.
29. M. Garner, A. Harvey, H. Johnson, *Int. Rev. Law Econ.* **62**, 105881 (2020).
30. D. A. Sklansky, *J. Crim. Law Criminol.* **96**, 1225 (2005).
31. ACLU, "Stop and Frisk in Chicago," American Civil Liberties Union of Illinois, 2015.
32. J. Byrne, R. Pearson, J. Gorner, "Mayor Rahm Emanuel rejects President Donald Trump's call for Chicago police to use stop-and-frisk tactics," *Chicago Tribune*, 9 October 2018.
33. Police Accountability Task Force, "Recommendations for Reform: Restoring Trust Between the Chicago Police and the Communities They Serve," 2016.
34. E. McClelland, *Politico*, 29 September 2018.
35. WTTW, "Power, Politics, & Pride: Afro-American Patrolmen's League," https://interactive.wttw.com/dusable-to-obama/afro-american-patrolmens-league [accessed 28 January 2021].
36. M. M. Adams, "The Patrolmen's Revolt: Chicago Police and the Labor and Urban Crises of the Late Twentieth Century," thesis, University of California, Berkeley (2012).
37. B. A. Ba, "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago," Working paper, 2019.
38. N. Rim, R. G. Rivera, B. A. Ba, In-group Bias and the Police: Evidence from Award Nominations (2019). Faculty Scholarship at Penn Law. 2132. https://scholarship.law.upenn.edu/faculty_scholarship/2132.
39. P. Atiba Goff, K. Barsamian Kahn, *Soc. Issues Policy Rev.* **6**, 177–210 (2012).
40. B. Ba, D. Knox, J. Mummolo, R. Rivera, The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago [Source Code] (2021). https://doi.org/10.24433/CO.4519202.v1.

ACKNOWLEDGMENTS

Data for this project were gathered in collaboration with the Invisible Institute, with the help of S. Stecklow and E. Herman. We thank R. Ryley and the Lucy Parsons Lab for generously sharing data. We thank R. Austin, P. Bayer, S. Black, T. Burch, B. Canes-Wrone, T. Clark, J. Grimmer, A. Harris, D. Kronick, A. Harvey, H. Jefferson, G. Lenz, B. Magaloni, T. Mendelberg, V. Weaver, A. White, and attendees of the 2019 Princeton Policing Conference for helpful feedback. This project was approved by institutional review boards at Columbia University; Princeton University; University of California, Irvine; and University of Pennsylvania. **Funding: Author contributions:** All authors contributed equally to this article. **Competing interests:** D.K. and J.M. are paid consultants for the U.S. Department of Justice Civil Rights Division and the American Civil Liberties Union. R.R. worked for the Invisible Institute and the plaintiffs in *Campbell v. City of Chicago* between 2017 and 2018. **Data and materials availability:** All data, interactive replication materials, and guaranteed reproducible results are publicly available as a Code Ocean capsule (*40*). We encourage readers to probe and extend our analyses.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/696/suppl/DC1
Supplementary Text
Figs. S1 to S20
Tables S1 to S8
References (*41*–*46*)

17 July 2020; accepted 7 December 2020
10.1126/science.abd8694

Science

The role of officer race and gender in police-civilian interactions in Chicago

Bocar A. Ba, Dean Knox, Jonathan Mummolo and Roman Rivera

Science **371** (6530), 696-702.
DOI: 10.1126/science.abd8694

Diversity in policing

In the wake of high-profile police shootings of Black Americans, it is important to know whether the race and gender of officers and civilians affect their interactions. Ba *et al.* overcame previous data constraints and found that Hispanic and Black officers make far fewer stops and arrests and use force less than white officers, especially against Black civilians. These differences are largest in majority-Black neighborhoods in the city of Chicago (see the Perspective by Goff). Female officers also use less force than male officers. These effects are supportive of the efficacy of increasing diversity in police forces.

Science, this issue p. 696; see also p. 677

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

science.sciencemag.org/content/371/6530/696/suppl/DC1

Supplementary Materials for

The role of officer race and gender in police-civilian interactions in Chicago

Bocar A. Ba, Dean Knox*, Jonathan Mummolo*, Roman Rivera

*Corresponding author. Email: dcknox@upenn.edu (D.K.); jmummolo@princeton.edu (J.M.)

Published 12 February 2021, *Science* **371**, 696 (2021)
DOI: 10.1126/science.abd8694

This PDF file includes:

Supplementary Text
Figs. S1 to S20
Tables S1 to S8
References

Table of Contents

List of Figures

List of Tables

S1 Detailed Description of Data

S1.1 Coding Race and Ethnicity

We determine race/ethnicity of CPD officers based on demographic data obtained from the CPD through FOIA. The CPD usually classifies race/ethnicity in at most 7 mutually exclusive groups: White/Caucasian, White Hispanic, Black/African American, Black Hispanic, Asian/Pacific Islander, Native American/Native Alaskan, and unknown/missing. However, there are inconsistencies in how races and ethnicities are coded across files. For example, some files do not include "Black Hispanic" as a racial category (very few officers are ever classified as Black Hispanic), and some files contain outdated racial categories which we update to the best of our ability. For consistency, we classify "Hispanic" and "White Hispanic" as "Hispanic"; "Black" and "Black Hispanic" (rare cases) as "Black." "White" in our analysis refers to non-Hispanic White. If an officer has multiple races associated with them across different datasets, we aggregate by most common non-missing races.

For Census and American Community Survey data, we construct corresponding race categories as follows: any Hispanic individual is coded Hispanic; White and Black are comprised of individuals who are coded as not Hispanic and White (Black) alone.

S1.2 CPD Data

The administrative data from the CPD used in this study span multiple datasets collected in collaboration with the Invisible Institute, Sam Stecklow, and Emma Herman over the course of three years (2016-2019). We obtained these records from the Chicago Police Department or Chicago Department of Human Resources via Freedom of Information Act (FOIA) or through court ordered releases stemming from requests made by Invisible Institute and Jaime Kalven. CPD provided the following data: rosters of all available current and past officers up to 2018, unit history data for individual officers from the 1930s to 2016, Tactical Response Reports from 2004 to 2018 (i.e. use of force reports), and arrest data with arresting officers and arrestee demographic information from 2001 to 2017. The Chicago Department of Human Resources provided data on officers' language skills up to 2019. We supplement our core data with data on "Stop, Question and Frisk" (SQF) activity between 2012-2015, which was shared by the Lucy Parson's Lab. Finally, the Automated Daily Attendance and Assignment sheet data for each police district between 2012 and 2015 was obtained via a FOIA request to the CPD and shared by Rachel Ryley.

These data and others have been used to construct rich profiles of Chicago Police Officers. While no file contains a unique identifier (star numbers change over time, names are common, etc.), we constructed unique officer profiles through a successive merge process described here. Each file contains some identifying information such as of demographic data (birth year, race, gender) or other characteristics (name, start/badge number, appointed date, resignation date, current unit). We used these identifying characteristics to first de-duplicate officers within a file and to then merge to pre-existing officer data with inter-file unique identifiers. The merging process itself is an iterative-pairwise matching method, where the officers in each dataset are repeatedly merged on identifying characteristics and any successful 1-to-1 match in a round removes the matched officers from the next round of merging.

The resulting data contains records on 33,645 police officers appointed between March of 1936 to February of 2018. The number of years and officers varies across analyses in our paper due to missing data (for example, assignment data only exists for the years 2012–2015).

S1.3 U.S. Census Merge

District and beat demographic data was constructed using the 2010 US Census data and the CPD's pre-2012 beat map. The centroid of each census tract was identified, then the demographic information of all the centroids inside a beat were aggregated to determine the beat's population and demographic composition. District demographics were determined by aggregating across all beats within that district. Post-2012 district and beat demographics were constructed based on the pre-2012 beat data discussed previously and using a crosswalk that maps pre-2012 beats to current (2018) beats and their respective districts.

S1.4 Preprocessing of Patrol Assignments

We restrict analysis to patrol assignments in which Black, Hispanic, or White officers serve. Asian/Pacific Islander and Native American/Alaskan Native officers are not examined due to small sample sizes. Within this subset, we further drop non-standard assignments (notably including "protest detail," "station supervisor," and "station security" assignments, as well as special assignments for training, compensatory time, and excused sick leave). Patrol assignments in which officers are indicated as non-present are also dropped. These steps are intended to ensure that officers nominally patrolling a beat are in fact actively circulating in the assigned geographic area, improving the plausibility of the common-circumstances assumption. For the same reason, we drop double shifts (patrol assignment slots in which the assigned officer served for more than one shift on the same day) to address the possibility that officers behave differently due to fatigue in these circumstances. We also eliminate officers assigned to non-standard watches (i.e., other than first through third watches). Finally, we drop officers at ranks other than "police officer." This step eliminates police sergeants, who serve in 8% of beat assignments but make very few stops and arrests, as well as legal officers, helicopter pilots, explosives technicians, and canine handlers.

S1.5 Preprocessing of Police Behavior Data

Events are merged to the remaining patrol assignments based on officer ID and date. This step discards a large number of events, including those involving officers of higher ranks and incidents occurring on rest days. For stops, arrests, and uses of force, we drop all events that occur outside of the reported patrol start/stop times, eliminating off-duty activity. Non-standard shifts are dropped (retaining only first–third watches), as are absences, shifts with assigned special duties (e.g. protest detail, station security, training), and double- or triple-duty days in which a single officer serves multiple shifts.

Stops for "dispersal" and "gang and narcotics-related loitering" are coded as loitering stops; those that are "gang / narcotics related" are coded as drug stops; "investigatory stops" and stops of "suspicious persons" are coded as suspicious behavior; and stops under the "Repeat Offender Geographic Urban Enforcement Strategy (ROGUES)" program are combined with the "other" category. For stops, if a single officer is reported as both primary and secondary stopping officer, only one event is retained.

Arrests for municipal code violations and outstanding warrants are categorized as "other."

S1.6 Patrol Assignment Process

During the period of our behavioral analysis, 2012-2015, CPD officers are assigned to nearly 6,000 distinct patrol tasks, each represented by a beat codes. The majority of these patrol tasks (about 51%) correspond to known geographic boundaries that average less than one square mile. For example, area 1431 corresponds to a known collection of city blocks in the CPD's Shakespeare district, to which over 6,000 officer-shift slots were assigned. Within this geographic area, our patrol assignment data distinguishes between standard patrol tasks (indicated with beat code "1431," roughly 4,000 officer-shifts) and additional patrol tasks distinguished by a alphabetical suffix (in this case, beat code "1431R," roughly 2,000 officer-shifts). In this case, beat code “1431” is assigned for officers working shifts 2 and 3, which overlap in the afternoon but not in the early morning. Beat code “1431R“ indicates a relief assignment and is always assigned for shift 1; it overlaps with the end of shift 3 and the beginning of shift 1 to ensure no gap in service exists. Other suffixes such as A or E refer to squads that rotate based on the operations calendar, but the numeric beat that precedes the suffix refers to the same area. Comparing officers within beat codes (and not generalizing to numeric beats) ensures that we are robustly controlling for the function officers perform when assigned to a specific beat code in addition to their geographic location.

Other beat codes correspond to fixed geographic areas not depicted in official CPD documents. For example, area 1442 is a nearby collection of city blocks in Shakespeare district that shares administrative tasks (e.g. community meetings (*41*) with area 1431); it is not indicated on any known CPD map and is allocated a smaller number of officer-shifts (roughly 1,000). In addition to these geographically demarcated patrol tasks, another 8.6% of beat codes correspond to desk duty in various parts of the city. These are indicated with “02” suffixes, such as assignment “1402,” which is subdivided further into desk assignments “1402A,” “1402B,” “1402C,” and so on. Throughout our analysis, we compare officers assigned to the same patrol task by exactly matching on patrol assignments, including both numeric codes and alphabetical suffixes. Though we cannot always pinpoint the geographic locations of beat assignments, this hyper-granular assignment data makes it highly plausible to assume that officers working under the same beat code are assigned to face common circumstances.

We now describe temporal variation in the specific officers that fill a patrol slot—the exogenous variation in officer identities (and hence demographics) that our analytic approach exploits. Officers request vacant “watches” including day, swing, and night shifts; these groups are allocated in a process based on seniority and the needs of the police unit. Officers also bid in advance for furlough assignments (leave days in excess of the usual days off each week) following a similar seniority-based process (*42*). Among officers available to serve in a patrol assignment—as determined by the watch allocation process and, most likely, additional consultation with unit commanders—specific daily patrol assignments are allocated based on predetermined rotating leave schedules. Officers are assigned to "day-off groups" in advance, which determine their non-furlough leave days according to a CPD-wide operations calendar issued late in the preceding year. For officers working standard 8.5-hour shifts, a duty cycle typically consists of six on-duty days, followed by two days off. (However, when the days off coincide with a weekend, officjers receive a third day off.) An example of an operations calendar is provided in SI Figure S11 along with a detailed description (*43*).

A notable feature of this system is that cycles do not occur on a weekly basis, so that most officers working the first shift on the first Tuesday of January will receive leave on some other Tuesday in January. Moreover, because both operations calendars and watch, furlough, and day-off group selections are made far in advance, officers effectively have no ability to anticipate fluctuations in civilian behavior encountered while on any particular patrol, *conditional on assigned beat and shift time*. Similarly, civilians do not have information on the officers they may encounter on a particular day and time. This assignment process therefore provides exogenous variation in patrol assignments that are orthogonal to conditions on the ground, allowing us to estimate the effect of deploying officers from one group vs. another while holding environmental conditions constant. See Section S3.1 for additional discussion of this point.

S2 Descriptive Analysis of CPD Composition and Assignment Patterns

S2.1 CPD Composition

Our newly acquired records allow us to reconstruct the history of CPD diversification over a much longer period than previously possible. While the CPD intermittently publishes annual reports with aggregate demographics, these data cover only 1995–2010 and 2016–2017. We extend the time-series backward to 1970, allowing for a comprehensive descriptive portrait of the evolution of the demographic correspondence between CPD personnel and city residents.

S2.2 District Characteristics

The CPD currently subdivides Chicago into 22 policing districts which correspond to CPD units, in which the majority of police officers work. A typical district covers roughly ten square miles. There were 25 districts (numbered 1–25) until 2012, at which time 3 smaller districts—ranking 18th, 21st, and 25th in land area (*44*)—were eliminated and merged with other districts. Districts 23 and 21 and District 13 were eliminated and absorbed into neighboring districts in March and December of 2012, respectively. While District 23 was mostly absorbed by District 19 and most of District 13 was absorbed by District 12, significant parts of District 21 were absorbed by Districts 1, 2, and 9.

Figure S3 illustrates the types of districts to which officers of each demographic group are assigned. This analysis takes each unique combination of racial/ethnic and gender, identifies all officers in that group, and then compute their assigned districts' average characteristics. Four dimensions are examined: violent crime rates, property crime rates, police officer density, and proportion of co-racial residents.

We now turn to two district-level analyses. Figure S4 plots the relationship between a police district's resident demographic profile (e.g. the proportion of residents that are Black) and officer demographic profile (the proportion of officers assigned to that district that are Black). White-dominated districts have virtually no minority officers assigned, and districts with sizeable minority populations tend to have more officers of the corresponding race. However, officers are disproportionately White compared to district residents: a number of districts dominated by Black residents nonetheless have sizeable contingents of White officers. For example, Wentworth (CPD District 2, depicted in Figure 1) is 95% Black, but 20% of officers assigned there are White. The disparity is even starker in Austin (CPD District 15), where a 93% Black resident population is policed by a unit that is 55% White. (See SI Section S1.3 for details on the computation of resident demographics.)

Figure S5 displays significant over-time changes in the racial composition of officers assigned to a district. In this figure, the vertical slice at 2010 corresponds to the results plotted in Figure S4. The proportion of Black officers assigned to some districts (e.g. districts 3, 5, 6, 7) while holding steady in others. Temporal discontinuities are due to changes in district boundaries or elimination of police districts.

S2.3 Officer Demographics and Patrol Assignments

Among officers assigned to a particular police district, considerable variation exists in the exact patrol assignments that officers receive. We examine each unit individually, tabulating officer race and shift time assignments (first, second and third watch, respectively corresponding to the nominal duty periods of midnight to 8 a.m., 8 a.m. to 4 p.m., and 4 p.m. to midnight). Figures S9–S10 depict the frequency of each shift period, respectively showing that the pattern of assignments differs dramatically by officer race and gender. For example, White officers in Wentworth (District 2) almost exclusively serve from 4 p.m. to midnight, whereas Black officers are more likely to be assigned to mid-day shifts.

Figures S6–S7 examine the pattern of patrol beat assignments by race/ethnicity and gender, respectively. They show that, for example, relative to White officers, Black officers are far more frequently deployed to assigned beat 202—which roughly corresponds to a patrol area in the district's southwest corner (depicted in Figure 1) that has extremely high police activity and a high concentration of Black residents. These results undermine analyses in a wide array of previous studies that aggregate at high levels of geography (for example, controlling for district or unit assignment) and which assume that officers face homogeneous conditions within these crude groupings.

S3 Officer Behavior

S3.1 Estimand

At a high level, the goal of our analysis is to evaluate the policy effect of a personnel reform that increases the representation of minorities in the CPD by assigning them to positions that would otherwise be filled by White individuals. The analysis is conducted at the level of the patrol assignment slot. Commanding officers are assumed to have a fixed set of patrol assignments that must be filled, where each slot is associated with a beat assignment and shift time (temporal window). Multiple slots may be available for a particular beat and shift time, but each slot can be filled by only one officer. We organize beat assignments into groups, indexed by i, based on unique combinations of month (M_i), day of week (D_i), shift time (first/second/third watch, S_i), and beat (B_i), or unique MDSBs.

A unit commander's deployment decisions—the allocation of an officer to a slot within an MDSB—are largely guided by (i) day-off group rotations, as determined by CPD operations calendars; and (ii) the set of available officers, determined in advance by officers and commanders based on typical conditions encountered in that MDSB. In SI Figure S12, we consider the hypothetical manipulation in which commanders choose to deploy a randomly chosen member of one group (e.g. available Black officers) in lieu of another group (e.g. available White officers). This ideal experiment is closely approximated by quasi-experimental rotation in the CPD operations calendar, which systematically removes officers (i.e., those on weekly leave) from the field in ways that are highly unlikely to correlate with unobserved fluctuations in crime levels or other local conditions. Because officers are deployed in their entirety, not net of a particular trait, this hypothetical decision by unit commanders means that the individual sent into the field carries a number of other characteristics—e.g., dialect, work experience, or home neighborhood—along with their group membership. Our analytic approach does not attempt to parse out these mechanisms, but instead seeks to estimate the difference in enforcement activities that unit commanders can expect from this hypothetical deployment policy change. SI Figure S12 further illustrates an important distinction between deployment decisions of unit commanders and the discretionary actions of patrol officers (choice of partner, walking patterns while on patrol, etc.). Because the latter decisions may be influenced by factors that are not captured by MDSB fixed effects, analyst should be cautious in controlling for these decisions to avoid introducing collider bias.

Patrol assignment slots within a MDSB are indexed by j. For each slot, the realized pattern of officer behavior is denoted $Y_{i,j}(R_{i,j})$, where $R_{i,j}$ is the demographic profile (race/ethnicity and/or gender) of the officer assigned to a particular slot. Our notation implicitly makes the stable unit treatment value assumption (*45*),which requires that (1) there do not exist finer gradations of officer identity (i.e., within the broad racial/ethnic and gender categories used) that would result in differing potential officer behavior, and (2) that potential outcomes do not very depending on the racial/ethnic and gender identities of officers assigned to other slots. (We explore the validity of this second assumption to the extent possible in SI Appendix S3.9, in which stops made by two officers are reanalyzed. In this section, we re-compute our estimates of differential stopping behavior after excluding the second reporting officer from our analysis; the resulting estimates are highly similar.)

The slot-level policy effect is the difference in potential outcomes (*46*) $Y_{i,j}(r) - Y_{i,j}(r')$, the change in behavior that would have realized if an officer of demographic profile r had been assigned to the patrol assignment slot, rather than another officer of profile r'. These slot-level counterfactual differences are fundamentally unobservable. Instead, we target the average policy effect within the subset of F MDSBs for which policy effects can be feasibly estimated (i.e., for which variation in officer demographic profiles exists). This quantity is

$$\delta = \frac{1}{F\overline{A}_i} \sum_{i=1}^{F} \sum_{j=1}^{A_i} Y_{i,j}(r) - Y_{i,j}(r'),$$

where A_i is the number of patrol assignment slots available within MDSB i and $\overline{A}_i$ is the average slot count across MDSBs. This can be rewritten as the weighted average of MDSB-specific effects, δ_i, with weights given by A_i.

$$\begin{aligned}
\delta &= \sum_{i=1}^{F} \frac{A_i}{\sum_{i'=1}^{F} A_{i'}} \left(\frac{1}{A_i} \sum_{j=1}^{A_i} Y_{i,j}(r) - Y_{i,j}(r') \right) \\
&= \sum_{i=1}^{F} \frac{A_i}{\sum_{i'=1}^{F} A_{i'}} \delta_i.
\end{aligned}$$

As we discuss in Section S3.1, a key identifying assumption is that

$$Y_{i,j}(r), Y_{i,j}(r') \perp\!\!\!\perp R_i \mid M_i = m,\ D_i = d,\ S_i = s,\ B_i = b.$$

Informally, this requires that minority officers are not selectively assigned to slots within MDSBs, at least in ways that matter for potential officer behavior. (Hypothetically speaking, this independence condition could be achieved even without adjusting for MDSB if White and non-White officers were randomly assigned locations and times to patrol.)

Our primary results estimate this quantity with an ordinary least squares (OLS) regression of the form $\hat{Y}_{i,j} = \hat{\alpha}_i + \sum_r \hat{\beta}_r \mathbf{1}(R_{i,j} = r)$, where $\hat{\alpha}_i$ represents a fixed effect for MDSB i and $\hat{\beta}_r$ represent regression coefficients for each officer race. For the estimator $\hat{\beta}_r - \hat{\beta}_{r'}$ to unbiasedly recover the overall quantity of interest, δ, requires the additional assumption that MDSB-specific policy effects are homogeneous, or that $\delta_i = \delta_{\mathscr{R}}$ for all i. It is well known that when this assumption is violated, OLS recovers the weighted average of δ_is with weights corresponding to the variance of officer demographic profiles within strata. To allow for the possibility of non-homogeneous policy effects and other departures from our modeling assumptions, we therefore apply a number of alternative estimators, which are described in detail in SI Section S3.8. As we show in SI Figures S15–S17, these alternative results are virtually identical to our primary results.

S3.2 Potential Threats to Validity

For full transparency, we highlight a number of possible threats to the validity of our analysis given our analytic goal. Confounding factors in this scenario include all variables that correlate with officer race and/or gender (depending on the analysis) in ways that violate the common-circumstances assumption. An example would be if Black and White officers were assigned to the same beat and shift, but Black officers were ordered to stay in their patrol cars the entire time while White officers were allowed to freely roam the beat, meaning Black and White officers faced systematically different working conditions for reasons beyond their control. Another example would be if there are unobserved differences within a MDSB (e.g., a beat is more dangerous on one particular Tuesday evening in a month, perhaps due to a scheduled protest) and officers of one group are preferentially assigned to patrol according to those differences. We assess that confounding of this type is extremely rare, because MDSB are defined in such a fine-grained way that comparisons are made within groups of roughly four beat-shifts (e.g., all Tuesday evenings in January 2012 for beat 251).

However, because we are not seeking to identify the effect of race per se, other correlates of officer race which do not violate the common-circumstances assumption do not obstruct our ability to evaluate this counterfactual. Examples of these innocuous correlates include: (1) Black and White officers possessing different levels of education which in turn lead to differential enforcement; or (2) male and female officers choosing to focus on different corners of their beats once assigned in ways that influence policing outcomes. In the latter case, officers still were *assigned* to face common circumstances (our key identifying assumption) but chose to turn a blind eye to certain subsets of civilian behavior. These facets represent different mechanisms through which the policy intervention of interest affects police-civilian interactions, but would not bias estimates relating to officer *deployment*.

S3.3 Shift Duration

We consider the possibility that stops, arrests, and uses of force are driven by different amounts of time spent patrolling. Even among officers assigned to a particular shift time (a nominal eight-hour patrol period), minor variation exists in the precise start and end of the officer's duty time. Of the officer-shifts analyzed, 86% are 9 hours in duration, with 8.5- and 8-hour shifts making up an additional 8% and 5%, respectively. In fixed-effect regression analyses that compare officers within unique MDSB combinations, we estimate that shifts of Black officers are 0.007 hours shorter (roughly 0.1% shorter) than their White counterparts assigned to the same MDSB, and Hispanic officer shift durations are virtually identical to those of White officers. Because these differences are two orders of magnitude smaller than reported differences in behavior, patrol time disparities are unlikely to be a mechanism driving observed racial gaps in stops, arrests, and force.

S3.4 Variation in Explanatory Variable

In analyzing how policing behavior varies with officer demographic characteristics, we compare the recorded decisions of different officers facing the same set of circumstances. To do so, we examine 653,087 unique combinations of month, day of week, shift number, and beat (MDSBs). Of these, 571,927 MDSBs have more than one assigned officer, a requirement to make any within-MDSB comparison. Single-officer MDSBs can arise if, for example, a beat requires only one officer to patrol and officer schedules are stable (e.g., if one individual consistently serves all first watches on Mondays for the month). To make cross-group comparisons, we further require that different officer groups have served in the same MDSB.

There are 294,927 MDSBs that contain overlap between multiple assigned officer racial/ethnic groups (e.g., one Black officer and one White officer); 229,114 MDSBs contain overlap between both female and male officers; and 52,562 MDSBs contain overlap between Spanish-speaking and non-Spanish-speaking Hispanic officers. Due to the smaller number of Hispanic officers and the resulting low overlap rates, power for detecting differences between Spanish- and non-Spanish-speaking Hispanic officer behavior is relatively low.

Out of all officers examined, 97% and 96% serve in at least one MDSB with feasible race and gender comparisons, respectively, and 92% of Hispanic officers serve in at least one MDSB with a feasible cross-language-group comparison. Of all MDSB examined, 45% and 35% contain feasible race and gender comparisons, respectively, and 18% of MDSBs with Hispanic officers contain a feasible cross-language-group comparison. The vast majority of these (82–88%, depending on the analysis) are between officers working at comparable but distinct times (e.g., first vs third Tuesday afternoons of a month); the remaining comparisons are between officers simultaneously assigned to a beat.

The primary driver of infeasibility appears to be the size of a patrol task (i.e., number of assigned officer-shifts). For example, MDSB with feasible racial comparisons contain an average of 5.8 officer-shifts, meaning that the chances of at least one non-White shift are much larger, whereas infeasible MDSB contain only 3.4 officer-shifts on average. Conversely, 21% of infeasible MDSB are in assignments with beat-code suffixes A–D, typically indicating fine-grained subdivision of assignments (i.e., officers are assigned to more specific tasks, meaning that fewer comparisons are available for any specific task); only 7% of feasible MDSB involve these beat-code suffixes. (Both p-values < 0.001.) Finally, feasibility of MDSB does not meaningfully vary with district majority racial composition; the proportions of infeasible MDSB in majority-Black, majority-Hispanic, majority-White, and no-majority districts are respectively 57%, 60%, 57%, and 59%. For illustrative purposes, Figure S8 plots the relative number of Black and White officers assigned to feasible and infeasible MDSB in each district type. The relatively symmetric distributions in Figure S8 alleviate concerns that within-MDSB effects are driven by, for example, one or two "unusual" White officers being compared to a much larger number of Black officers or vice versa.

S3.5 Main Results

Detailed results for Black-versus-White, Hispanic-versus-White, and female-versus-male officer deployment effects are presented in Figure S14 for all enforcement categories and subcategories. Tables S4–S6 respectively present numeric effect estimates for each of the above officer-group comparisons; Table S7 presents additional comparisons between Spanish-speaking and non-Spanish-speaking Hispanic officers. These tables also report Benjamini-Hochberg adjusted p-values for 92 hypotheses (23 outcome categories and subcategories, repeated for four officer-group contrasts) along with comparisons of effect sizes to typical enforcement volume for the reference group. Table S8 summarizes all numeric results and p-values.

S3.6 Design Comparison

To demonstrate how invalid analytic strategies can lead analysts to draw wildly incorrect conclusions about policing deployment, we compare results from our primary specification to those obtained by three alternative approaches that fail to achieve "apples-to-apples" comparisons due to two factors: (1) selection on the dependent variable and (2) the use of often-invalid comparisons between officers facing vastly differing contexts and pools of civilian behavior. The first approach we consider discards "zeroes," or patrol assignments in which the policing behavior under study does not occur, while still making valid within-MDSB comparisons. We find that discarding zeroes alone can lead analysts to mistakenly estimate large effects of the opposite sign. We then examine two additional invalid analytic approaches that discard successively more information: making a broader set of often-invalid comparisons between officers in the same district-month, ignoring large variation in the beat and shift contexts that each officer demographic receives; and making unrestricted comparisons between all officers in the CPD over all periods. The resulting (incorrect) estimates range from dramatic overstatements of the true effect to substantial understatements or even estimates of the opposite sign. These results suggest the need for caution and increased rigor in research design in the study of policing.

S3.7 Heterogeneity of Deployment Effects

Figure S19 demonstrates that differences in behavior between male and female officers remain even when comparing within members of a single racial/ethnic group. Female officers consistently use less force overall, and less force toward Black civilians in particular, than their co-racial/co-ethnic male counterparts. In addition, we find that Black female officers make slightly fewer drug stops than Black male officers, though this gap is difficult to interpret given the larger number of stops for miscellaneous other reasons.

Figure S20 demonstrates that when considering stop volume, the effect of deploying Black officers instead of White officers varies substantially depending on resident demographics and time of day. White officers make far more stops than Black officers in districts where residents are primarily Black, and this gap is magnified during nighttime shifts (first and third shifts, respectively covering midnight–8 a.m. and 4 p.m.–midnight). White officers also make substantially more stops than their in White-majority districts. However, this disparate enforcement in predominantly White-resident districts is most apparent during the daytime shift (when minority non-resident visitors are likely more numerous), rather than nighttime shifts.

S3.8 Alternative Estimators

Our primary analysis of officer behavior uses OLS regression with MDSB fixed effects, of the form $Y_{i,j} = \alpha_i + \sum_r \beta_r \mathbf{1}(R_{i,j} = r) + \varepsilon_{i,j}$, where α_i represents a fixed effect for MDSB i. As we discuss in SI Section S3.1, this estimator will deviate from the desired average policy effect (i.e., the average effect of replacing White officers assigned to a particular patrol assignment slot with a minority officer on resulting stop, arrest, and use-of-force volume) if MDSB-specific policy effects are highly variable in a way that is associated with the proportion of minority/female officers that are assigned to MDSBs (in this case, it is well known that OLS recovers the weighted average of MDSB-specific policy effects, where weights are determined by variance of officer race within the MDSB).

To gauge robustness of our results to the violation of this assumption, we present alternative estimates in SI Figures S15–S17 below. The first alternative estimator takes the within-MDSB difference in behavior between average patrol assignments between given officer demographic profiles, then aggregates these according to the number of patrol assignment slots in each MDSB. Following the notation defined in SI Section S3.1, this estimator can be written as

$$\sum_{i=1}^{F} \frac{A_i}{\sum_{i'=1}^{F} A_{i'}} \sum_{j=1}^{A_i} \left(\frac{\sum_{j=1}^{A_i} Y_{i,j} \mathbf{1}(D_{i,j} = d)}{\sum_{j=1}^{A_i} \mathbf{1}(D_{i,j} = d)} - \frac{\sum_{j=1}^{A_i} Y_{i,j}\ \mathbf{1}(D_{i,j} = d')}{\sum_{j=1}^{A_i} \mathbf{1}(D_{i,j} = d')} \right).$$

To assess the extent to which results are driven by large MDSBs, we further compute the unweighted average of MDSB-specific estimated effects:

$$\frac{1}{F} \sum_{i=1}^{F} \sum_{j=1}^{A_i} \left(\frac{\sum_{j=1}^{A_i} Y_{i,j}\ \mathbf{1}(D_{i,j} = d)}{\sum_{j=1}^{A_i} \mathbf{1}(D_{i,j} = d)} - \frac{\sum_{j=1}^{A_i} Y_{i,j}\ \mathbf{1}(D_{i,j} = d')}{\sum_{j=1}^{A_i} \mathbf{1}(D_{i,j} = d')} \right).$$

Finally, we consider the possibility that observed demographic differences in officer behavior are driven by differences in experience between officer groups. If this were the case, it would undermine the applicability of our results to the effect of a hiring reform that brought in additional minority rookie officers. To examine whether these differences impact our results, we extend the regression specification by adding additional linear and quadratic terms for each officer's length of service. Specifically, we estimate $Y_{i,j} = \alpha_i + \sum_r \beta_r \mathbf{1}(R_{i,j} = r) + \gamma_1 S_{i,j} + \gamma_2 S_{i,j}^2 + \varepsilon_{i,j}$, where $S_{i,j}$ is the length of service of officer j as of the month corresponding to MDSB i occurred. This robustness test generally does not alter our substantive conclusions;

the sole exception is that differences in the number of arrests made by Spanish-speaking and non-Spanish-speaking Hispanic officers vanish when adjusting for length of service.

Finally, we examine an alternative random effects estimator, in which individual officers are assumed to have average daily enforcement activity drawn from officer-race-specific normal distributions. We thank an anonymous reviewer for this suggestion. This modeling approach captures one particular form of dependence in an officer's enforcement activities: the possibility that all shifts will exhibit a higher average level of enforcement, relative to other officers. (In contrast, the block bootstrap approach of all other specifications ensures that inferences are robust to *arbitrary* within-officer dependence, including the following: overwork in one shift leading to less effort exerted in the following shift, life events leading to fluctuation in officer behavior on a timescale of a few months, or discontinuous life events like birth of a child leading to long-term changes in officer behavior.) We further note that the assumptions of the random-effects estimator are unlikely to be satisfied in the context of policing.

Empirically, we find that the alternative random-effects estimator is relatively consistent with our main specification. To the extent that results diverge, they tend to follow two broad patterns. First, point estimates are systematically anticonservative. Across all outcomes, 76% of racial deployment effect estimates from the random-effects model are larger than the corresponding estimates from our primary specification, ordinary least squares with MDSB-specific fixed effects. In the aggregate, random-effect estimates are 11 percentage points larger on average than fixed-effect estimates. We conjecture that these results are due to skew in officer behavior. Second, confidence intervals are also systematically anticonservative. The random-effects estimator produces 95% confidence intervals that are on average 43% shorter than the block-bootstrap confidence intervals of our primary fixed-effects specification. We conjecture this is because the random-effects estimator assumes away many forms of dependence between an officer's shifts. This behavior is illustrated in Figure S18.

S3.9 Robustness Checks: Multiple Stopping Officers

Data on stops of civilians indicate that the vast majority of enforcement is jointly conducted by two officers, but one is listed as the primary officer in police records. (Arrest records also often contain more than one officer, but contain have no indication of the lead officer.) In our main analysis, we treat a stop by two officers as two incidents in the data, as both officers contribute to the decision to engage a civilian. To gauge the extent to which this decision drives our results, we present an alternative analysis of stops in which we use only data on first reporting officers, respectively. Results are substantively unchanged. Note that the reduction in female-officer drug stops (versus male officers) loses significance, and female officers appear to take the lead in certain stop types. Specifically, female officers appear to be listed as first officer on more stops for traffic violations and miscellaneous other reasons, compared to male officers facing identical circumstances. (An apparent deployment effect on stops of black civilians appears to be marginally significant, but loses significance after multiple testing correction.) The reason for this gap in gender patterns is unknown. However, other general patterns remain substantively identical, with smaller coefficients reflecting the fact that roughly half of all stop events have been discarded. Given the lack of information on arresting officer roles, we do not conduct a similar robustness test for the arrest analysis. Our only option would be to drop an officer from each arrest at random, which would in expectation merely produce identical patterns with attenuated coefficients.

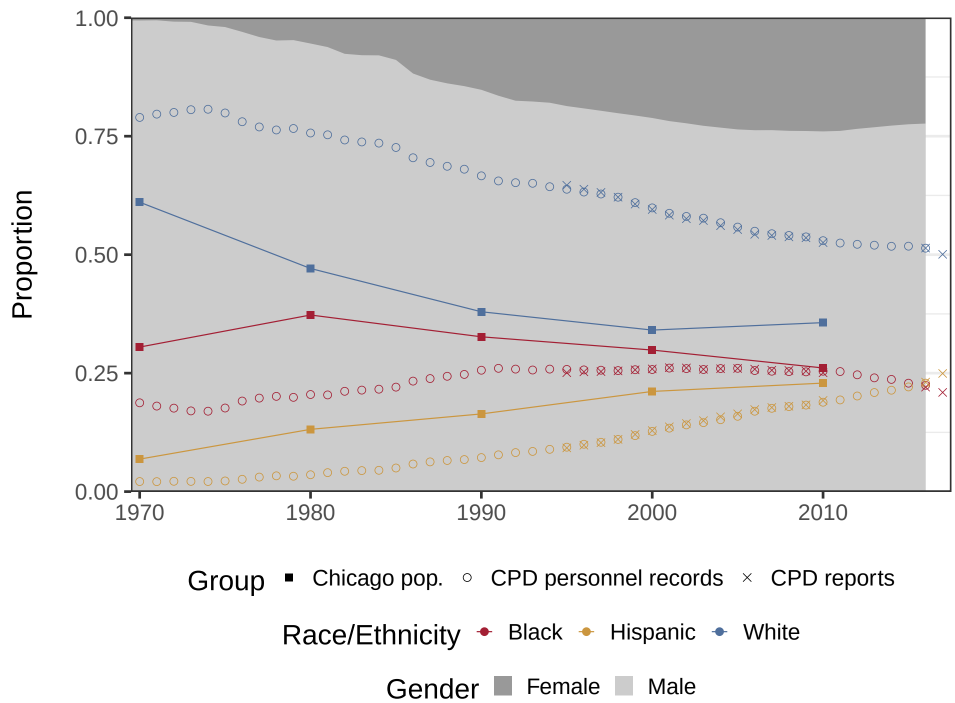

Figure S1. Composition of CPD officers and city residents over time. Red, yellow, and blue ○ depict the proportion of Black, Hispanic, and White active CPD officers in December of each year, according to our personnel records. Dark and light gray regions respectively indicate the proportion of female and male officers, using the same data. Data from CPD annual reports on the demographics of sworn and exempt/command officers are available only for 1995–2010, 2016 and 2017 (not shown); these are shown with ×. When available, these reports closely track our personnel data and increase confidence in our historical reconstructions. Lines indicating city of Chicago decennial Census proportions for each racial/ethnic group, tabulated by the National Historical Geographic Information System, are shown with ■ for reference.

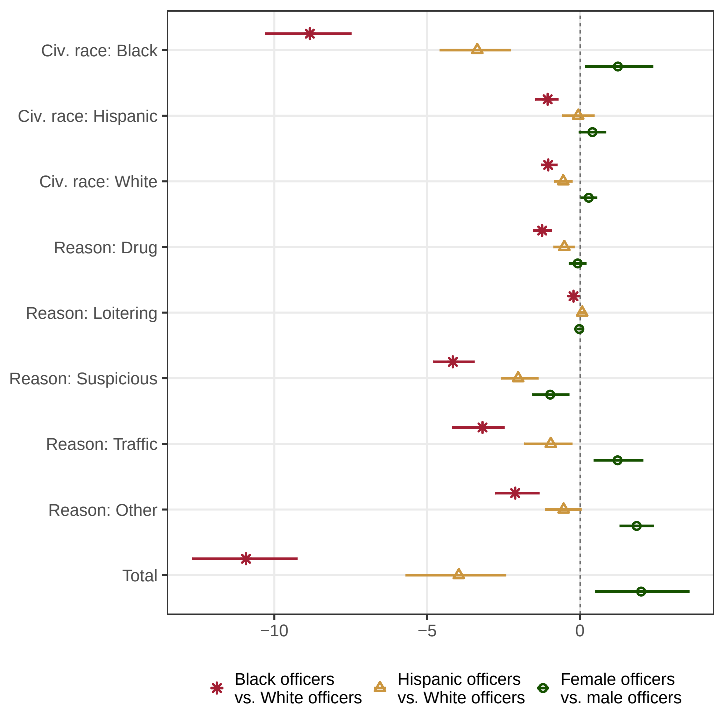

Figure S2. Alternative results using first-officer stops only. Relative to primary results examining all stops, the magnitude of deployment effects is smaller due to the exclusion of enforcement activities from analysis. However, substantive interpretations and patterns of statistical significance are generally unchanged, with the notable exception that female officers appear to make more stops than male officers facing identical circumstances. These gaps are driven by increased stops for traffic violations and miscellaneous other reasons.

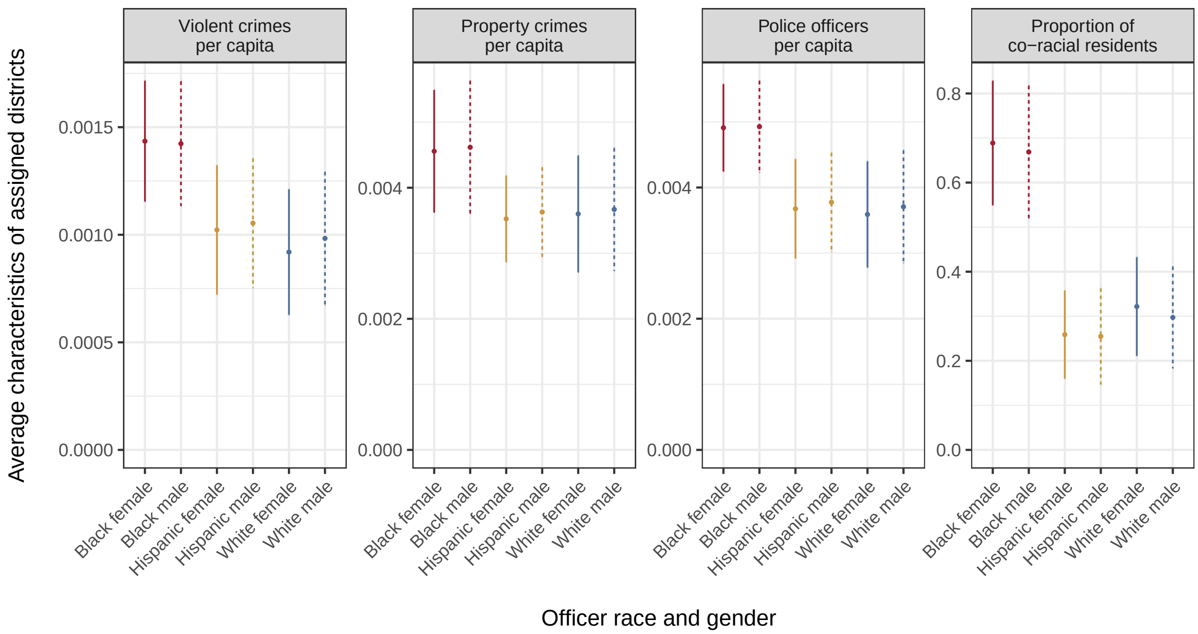

Figure S3. Average characteristics of assigned geographic districts. Results depicted for various officer groups, from 2006–2016. Confidence intervals cluster on district-month.

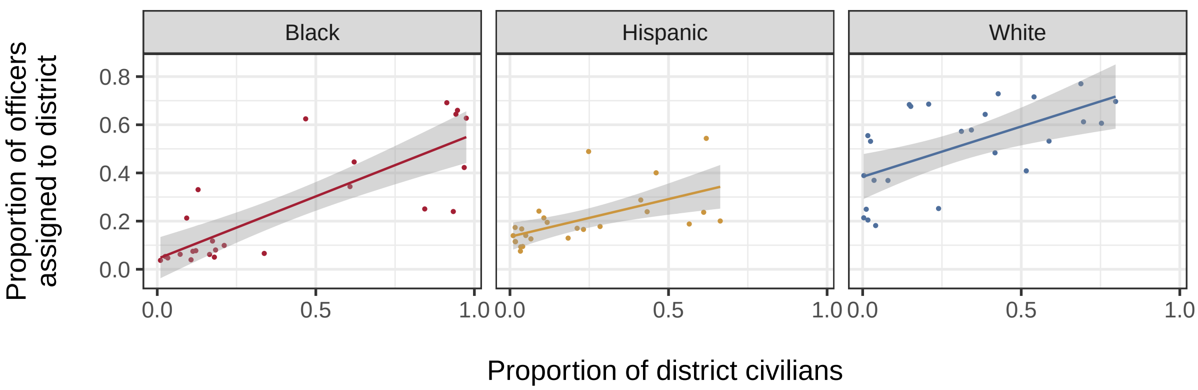

Figure S4. Racial and ethnic composition of officers' assigned districts. In each panel, each point represents a police district. The horizontal axis indicates the proportion of civilians of a given racial/ethnic group residing in 2010 Census data, and the vertical axis depicts the share of officers assigned to that district in January 2010 from the same racial/ethnic group.

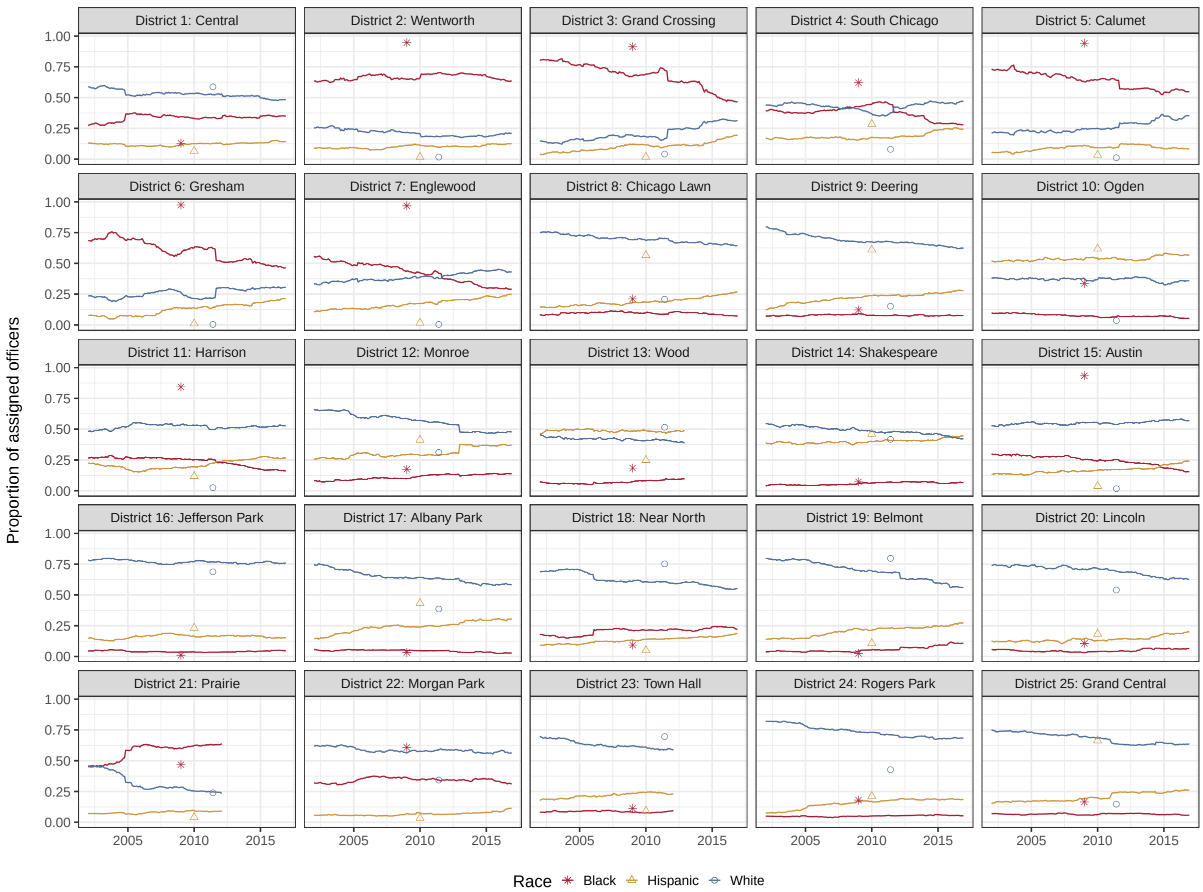

Figure S5. Racial composition of police districts. Each panel depicts a geographic police district. Points represent the racial composition of district residents. Lines represent monthly proportions of officers assigned to a district that belong to each racial group. Districts 21, 23, and 13 were eliminated during the observation period.

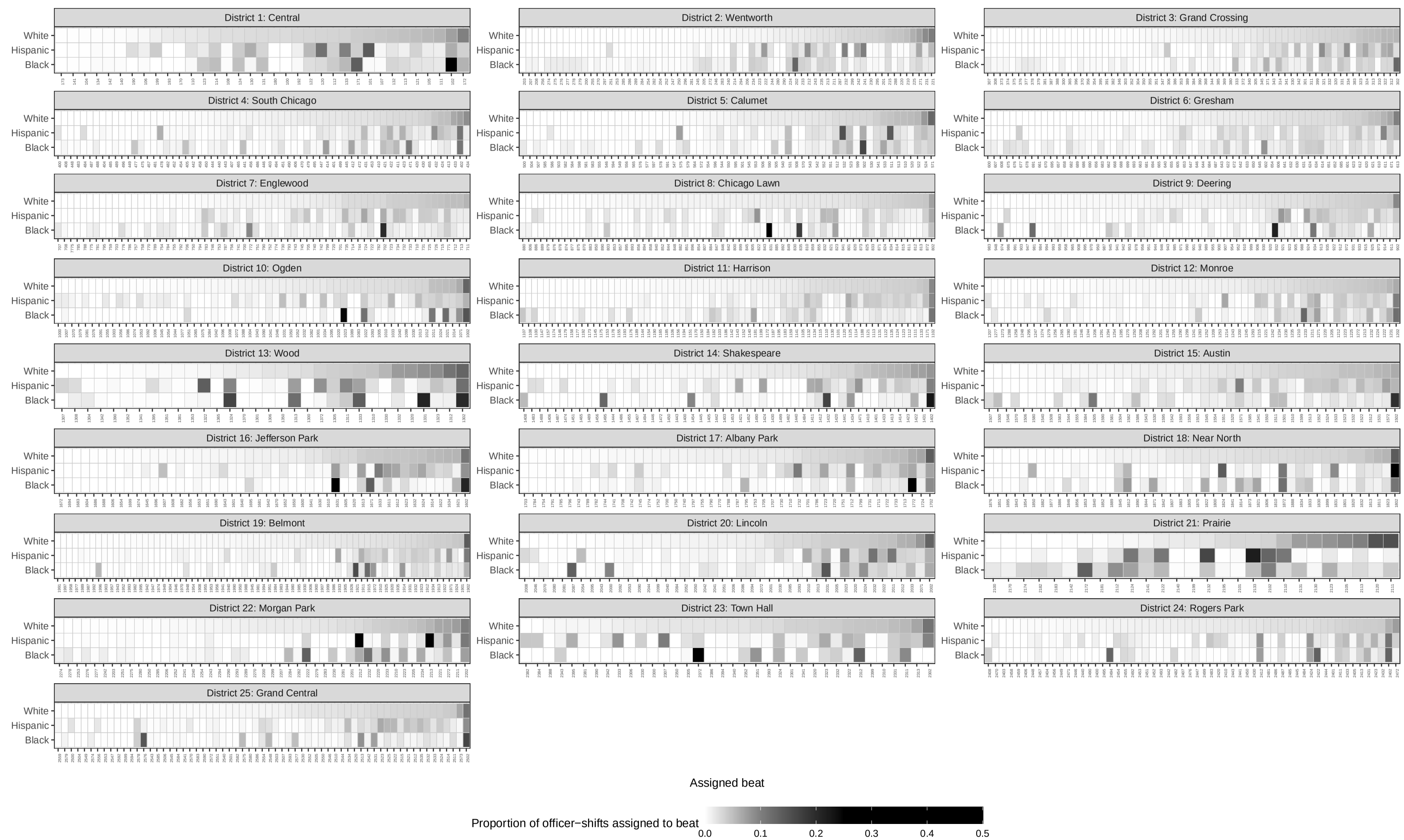

Figure S6. Assigned beat by officer race. Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for Black, Hispanic, or White officers to beats assignments. Darker cells indicate a higher proportion of assignments to a beat, and entries in a row sum to unity. The figure demonstrates that within any particular district, Black and Hispanic officers are called to serve at very different locations ($p < 0.001$ for all within-district statistical tests of independence).

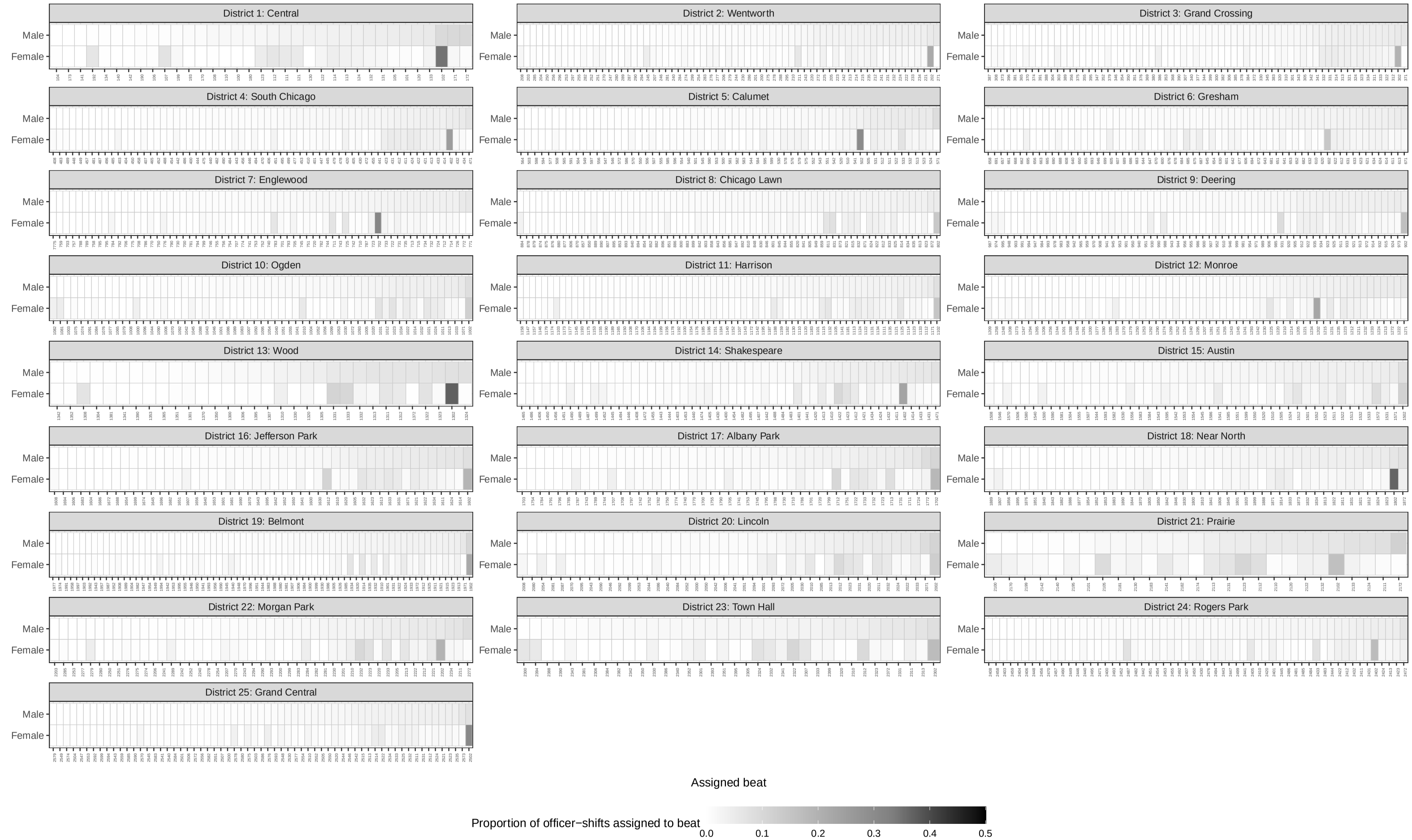

Figure S7. Assigned beat by officer gender. Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for female or male officers to beats assignments. Darker cells indicate a higher proportion of assignments to a beat, and entries in a row sum to unity. The figure demonstrates that within any particular district, female and male officers are called to serve at very different locations ($p < 0.001$ for all within-district statistical tests of independence).

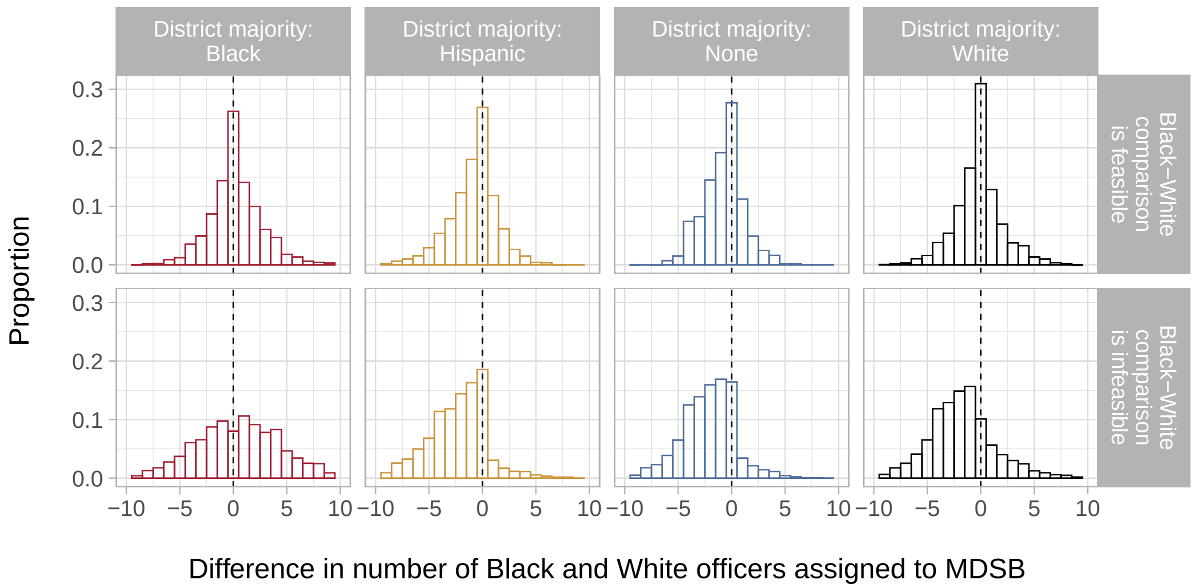

Figure S8. Relative number of Black and White officers assigned to MDSB. Among the feasible MDSB studied in our analyses, the relative number of assigned Black and White officers is well balanced and comparable all district types (Black, Hispanic, White, and no majority districts). This similarity demonstrates that our results are not driven by a small number of White officers in Black-majority areas. However, reasons for infeasibility of the Black-White officer comparison differ somewhat across district types (i.e., whether an MDSB is infeasible because no Black officers are assigned or because no White officers are assigned). In Black-majority districts, the two sources of infeasibility are roughly equal in frequency. However, in Hispanic-majority, no-majority, and (to a lesser extent) White-majority districts, MDSB are more likely to be infeasible because no Black officers are assigned. This is because there are simply fewer Black officers available to serve in these districts. This figure alleviates concerns that within-MDSB effects are driven by, for example, one or two "unusual" White officers being compared to a much larger number of Black officers or vice versa.

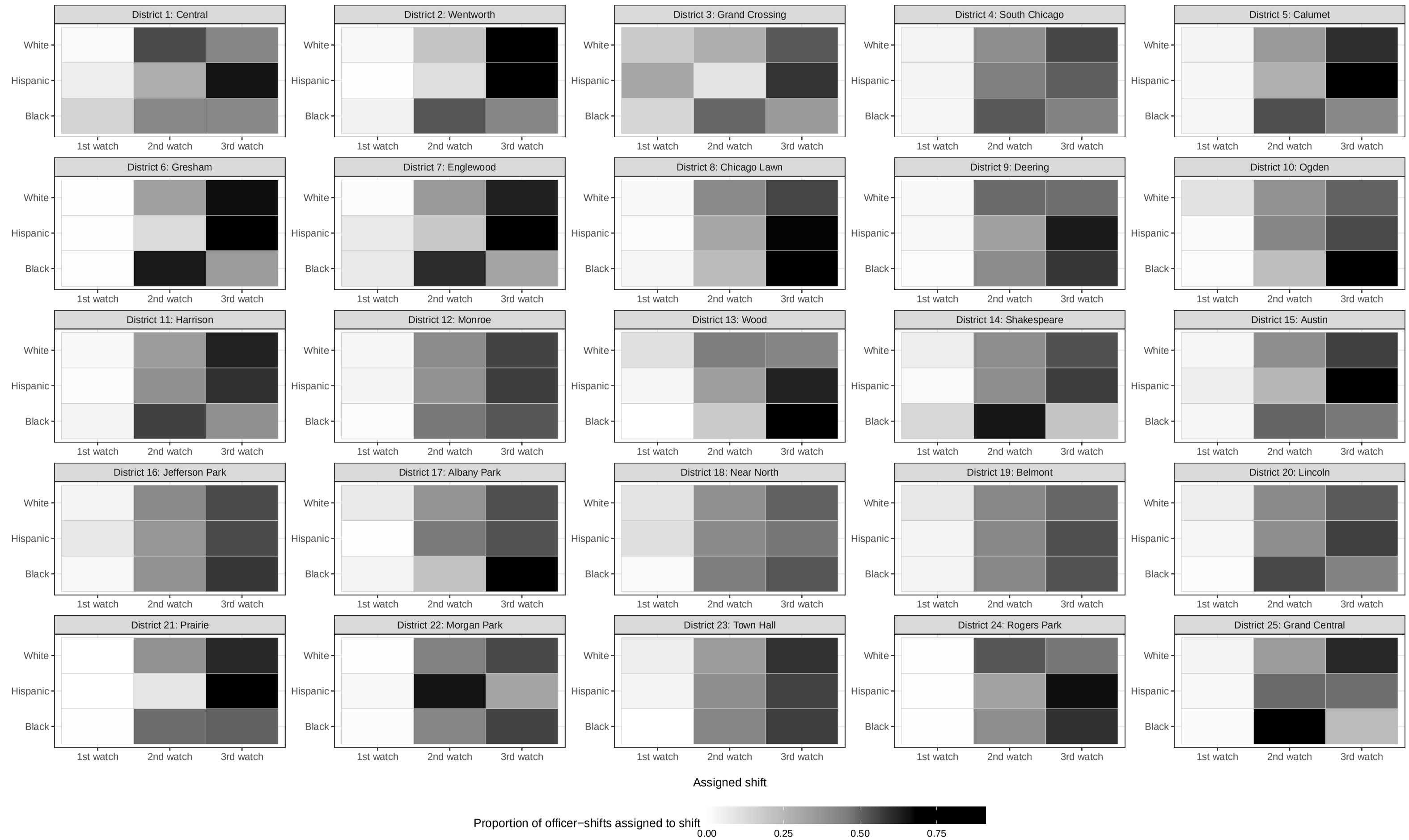

Figure S9. Assigned shift time by officer race. Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for Black, Hispanic, or White officers to first watch (midnight to 8 a.m.), second watch (8 a.m. to 4 p.m.), and third watch (4 p.m. to midnight). Darker cells indicate a higher proportion of assignments to a shift time, and entries in a row sum to unity. The figure demonstrates that within any particular district, Black and Hispanic officers are called to serve at very different times of day ($p < 0.001$ for all within-district statistical tests of independence).

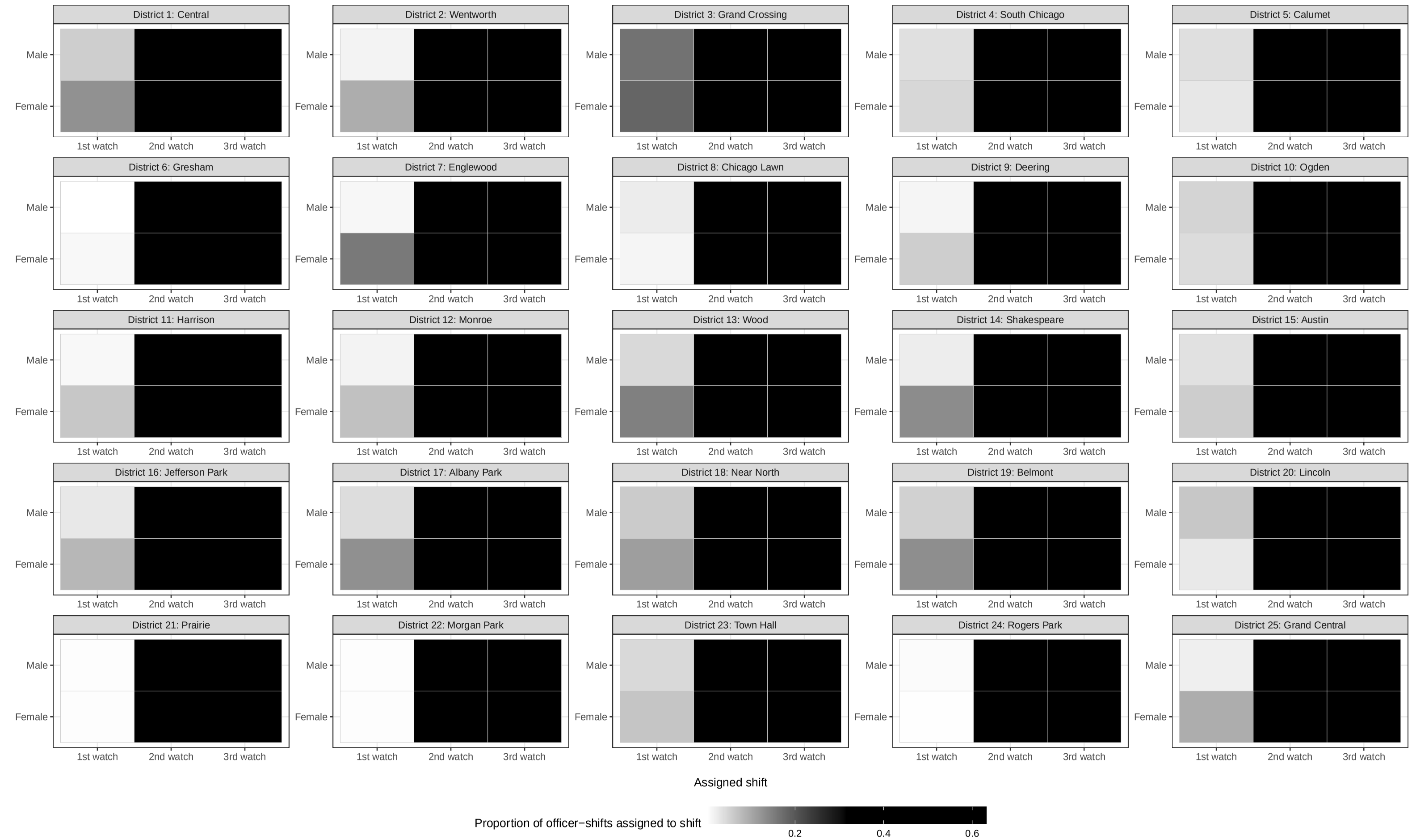

Figure S10. Assigned shift time by officer gender. Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for female or male officers to first watch (midnight to 8 a.m.), second watch (8 a.m. to 4 p.m.), and third watch (4 p.m. to midnight). Darker cells indicate a higher proportion of assignments to a shift time, and entries in a row sum to unity. The figure demonstrates that within any particular district, female and male officers are called to serve at very different times of day ($p < 0.001$ for all within-district statistical tests of independence).

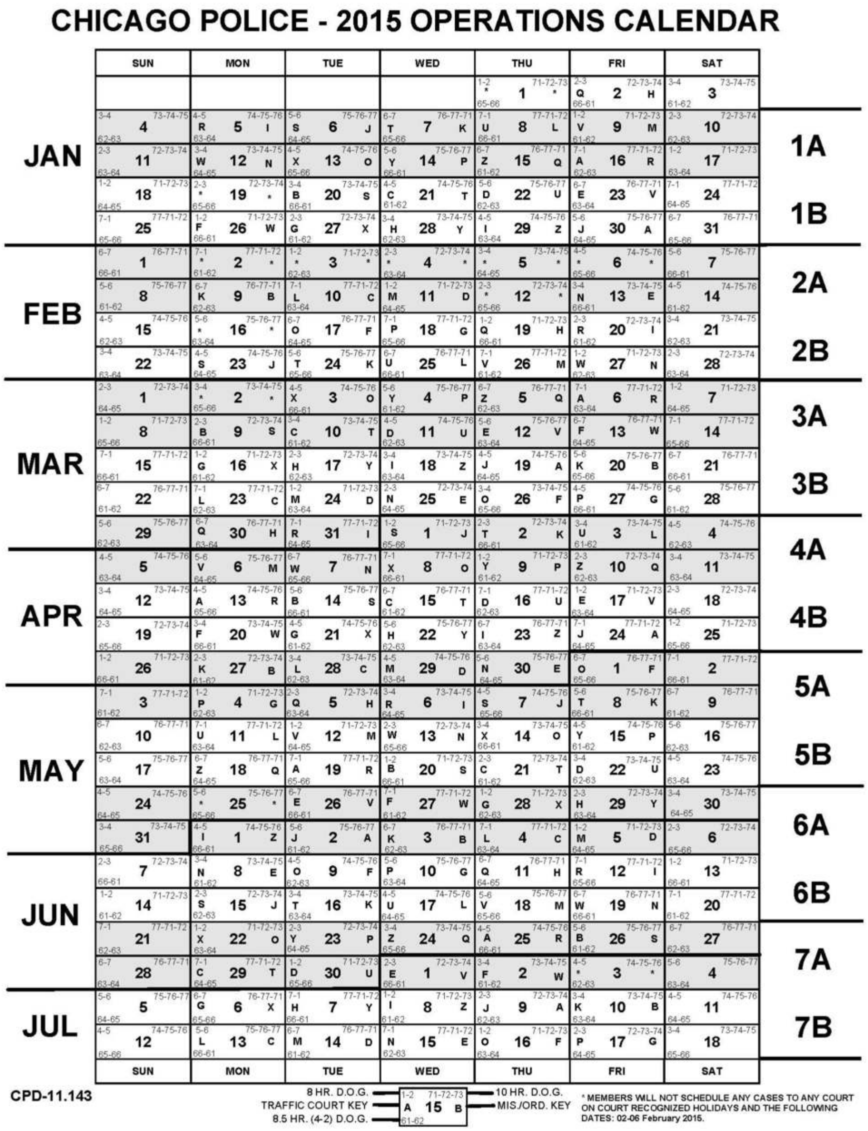

Figure S11. Officer availability for patrol assignments is determined by fixed operations calendars. The figure provides an example operations calendar (*43*). The upper-left corner of each date cell indicates the “day-off groups” receiving leave on that date (for officers on standard eight-hour shifts). For example, on 1 January 2015, day-off groups #1 and #2 received leave; on 2 January, day-off group #1 returned to patrolling, but #2 remained on leave. Officers typically work six days continuously, followed by two days off. (However, when leave coincides with a weekend, the leave is extended to three days.) These fixed rotation rules and preassigned groupings mean that in any particular patrol slot, the cycling of eligible officers provides an ignorable source of variation in patrol assignments, forming the basis of our identification strategy.

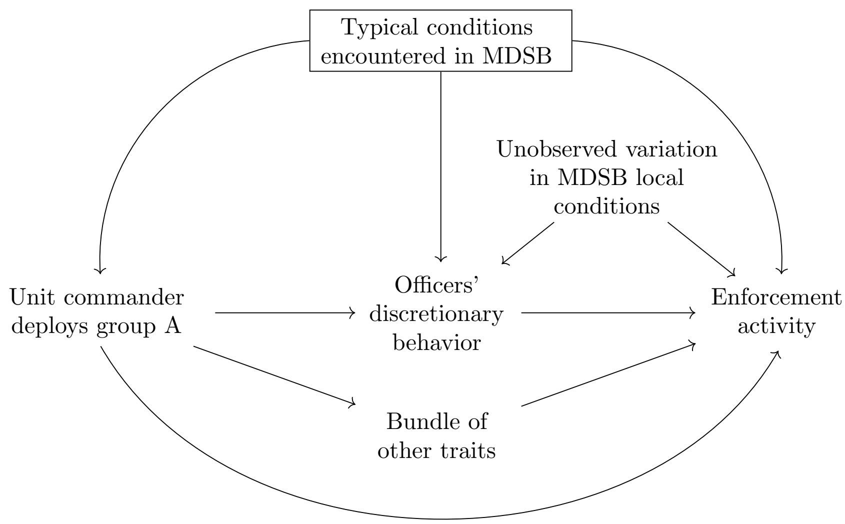

Figure S12. Data-generating process. Directed acyclic graph describing how a unit commander's decision to deploy a member of group A, as opposed to a member of group B, influences the enforcement activities recorded in policing data.

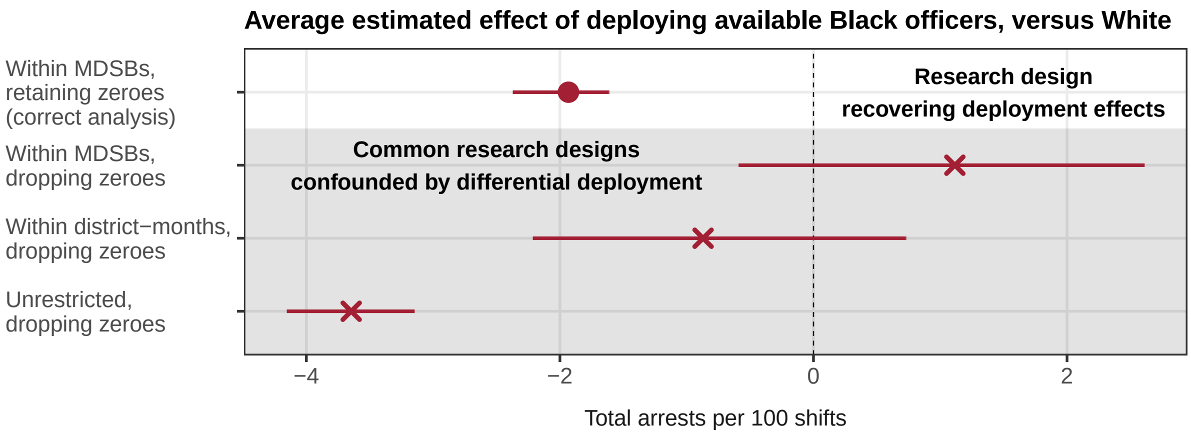

Figure S13. Comparison with common incorrect analytic approaches for estimating deployment effects. The figure displays the divergent estimates of deploying a Black officer relative to a White officer on arrests per 100 shifts that result from having the coarse of incomplete data available in most prior studies, i.e. either lacking micro data on the times and places of deployments (MDSBs), lacking data on officer shifts in which officers take no enforcement action, or both. The top estimate makes the correct comparison: comparing Black and White officers within the same MDSBs (i.e. facing similar circumstances), which yields a statistically significant decrease of about two arrests per 100 shifts. The second estimate from the top shows that by simply omitting shifts in which officers took no enforcement action, the point estimate changes from negative to positive, and the result is no longer statistically significant. Likewise, having no data on the times and places of deployments and lacking data on shifts with zero enforcement (the bottom estimate) yields a substantial overestimate of the effect.

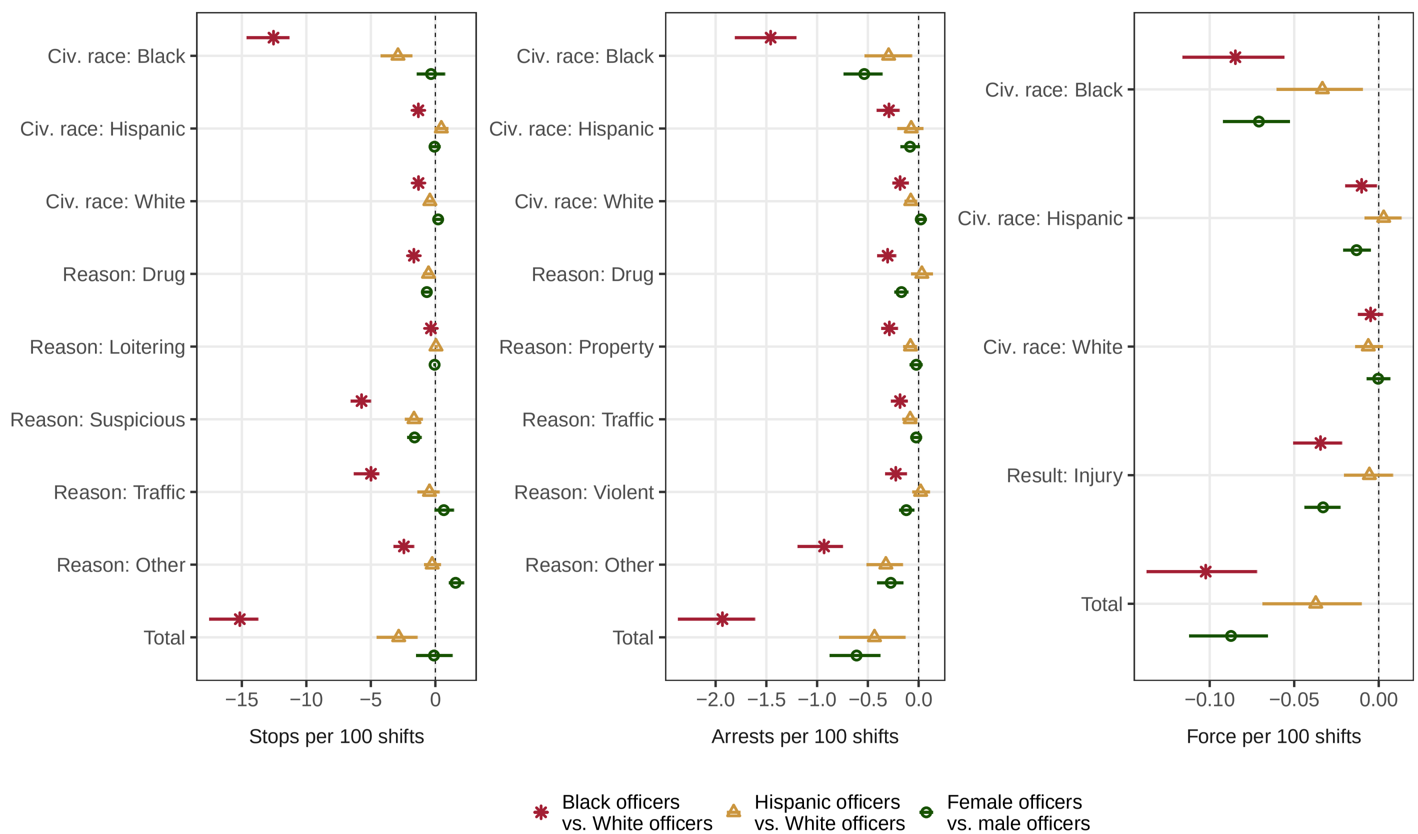

Figure S14. Behavioral differences across officer groups. The figure displays average within-MDSB differences between rates of stops, arrests and uses of force per 100 stops across various sub-categories of each outcome.

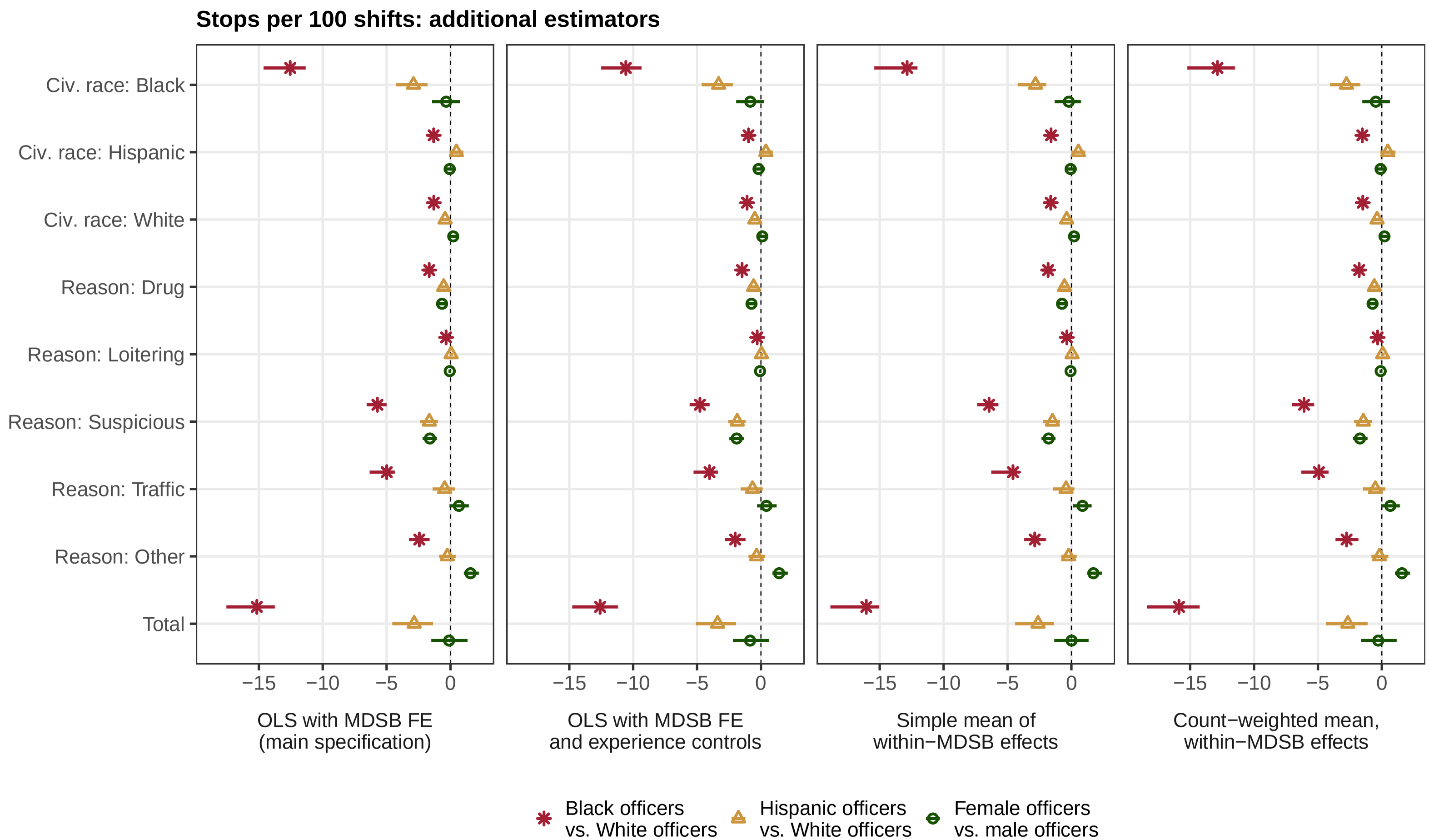

Figure S15. Stop outcomes, additional estimators. Points (error bars) depict estimated differences (95% block-bootstrap confidence intervals) in the number of civilian stops conducted by Black and Hispanic officers per shift, relative to White officers patrolling in the same month, day group (weekday/weekend), shift time (first/second/third watch), and beat. Results presented using four different estimators.

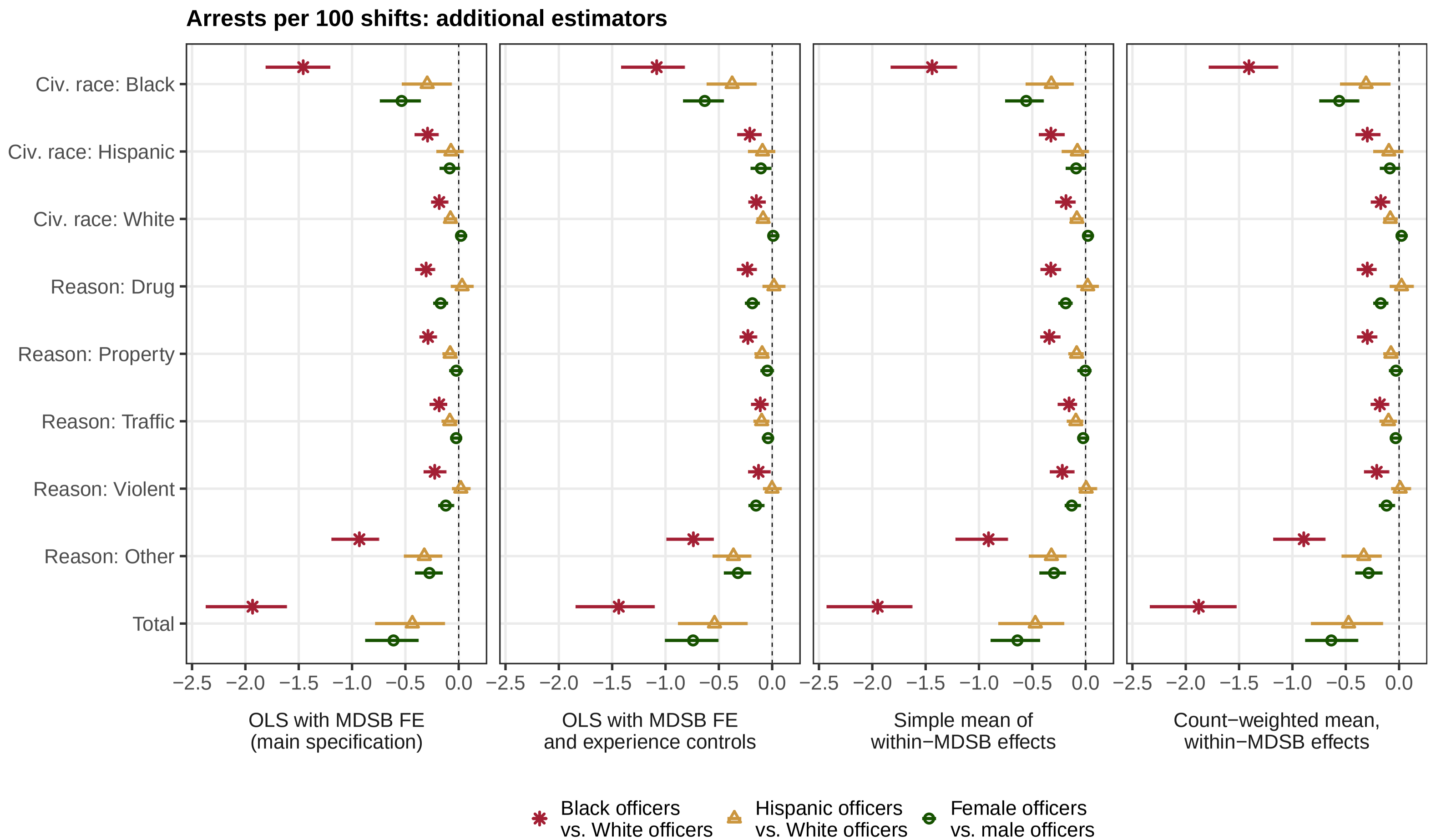

Figure S16. Arrest outcomes, additional estimators. Points (error bars) depict estimated differences (95% block-bootstrap confidence intervals) in the number of arrests conducted by Black and Hispanic officers per shift, relative to White officers patrolling in the same month, day group (weekday/weekend), shift time (first/second/third watch), and beat. Results presented using four different estimators.

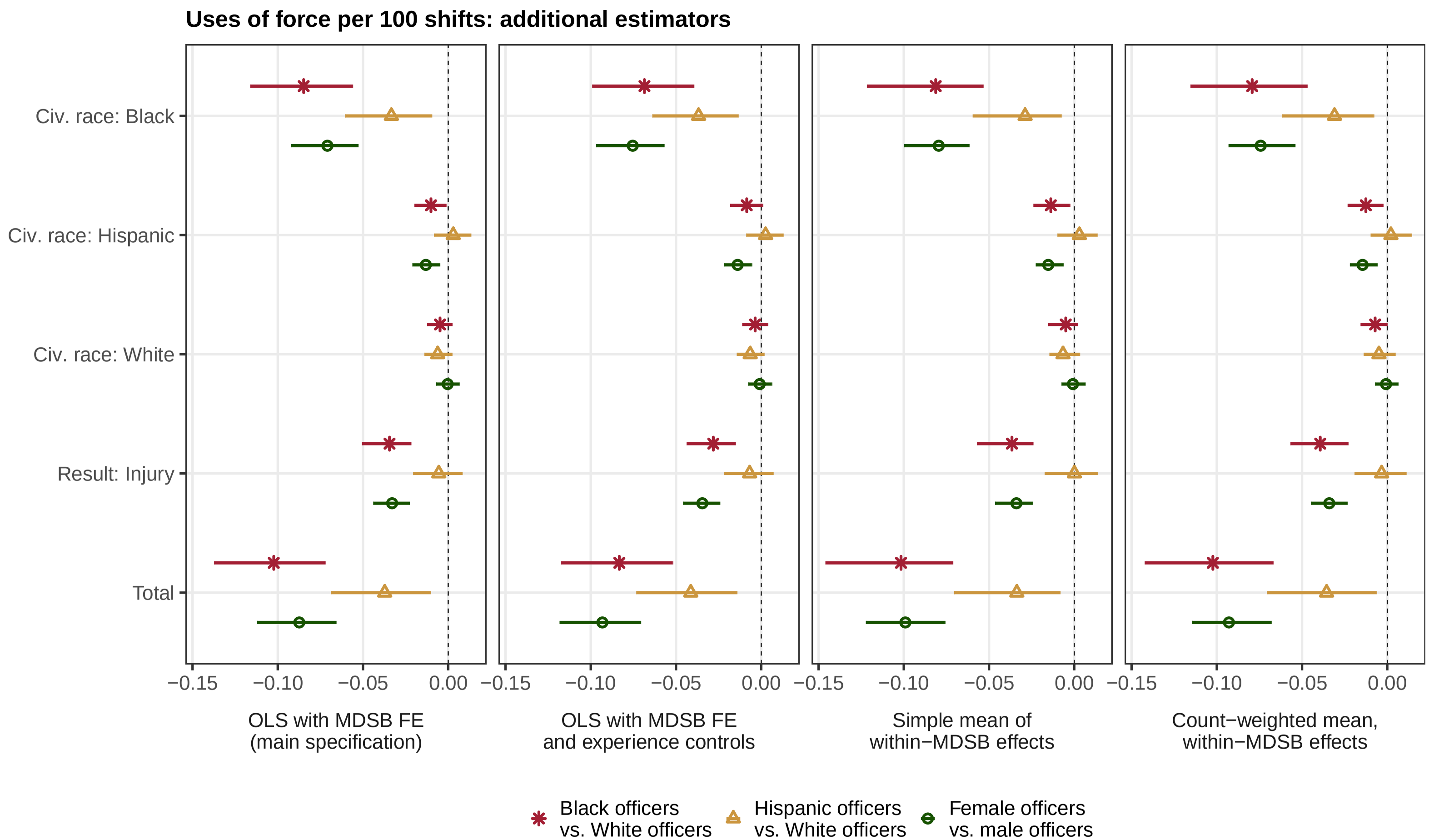

Figure S17. Use of force outcomes, additional estimators. Points (error bars) depict estimated differences (95% block-bootstrap confidence intervals) in the number of uses of force by Black and Hispanic officers per shift, relative to White officers patrolling in the same month, day group (weekday/weekend), shift time (first/second/third watch), and beat. Results presented using four different estimators.

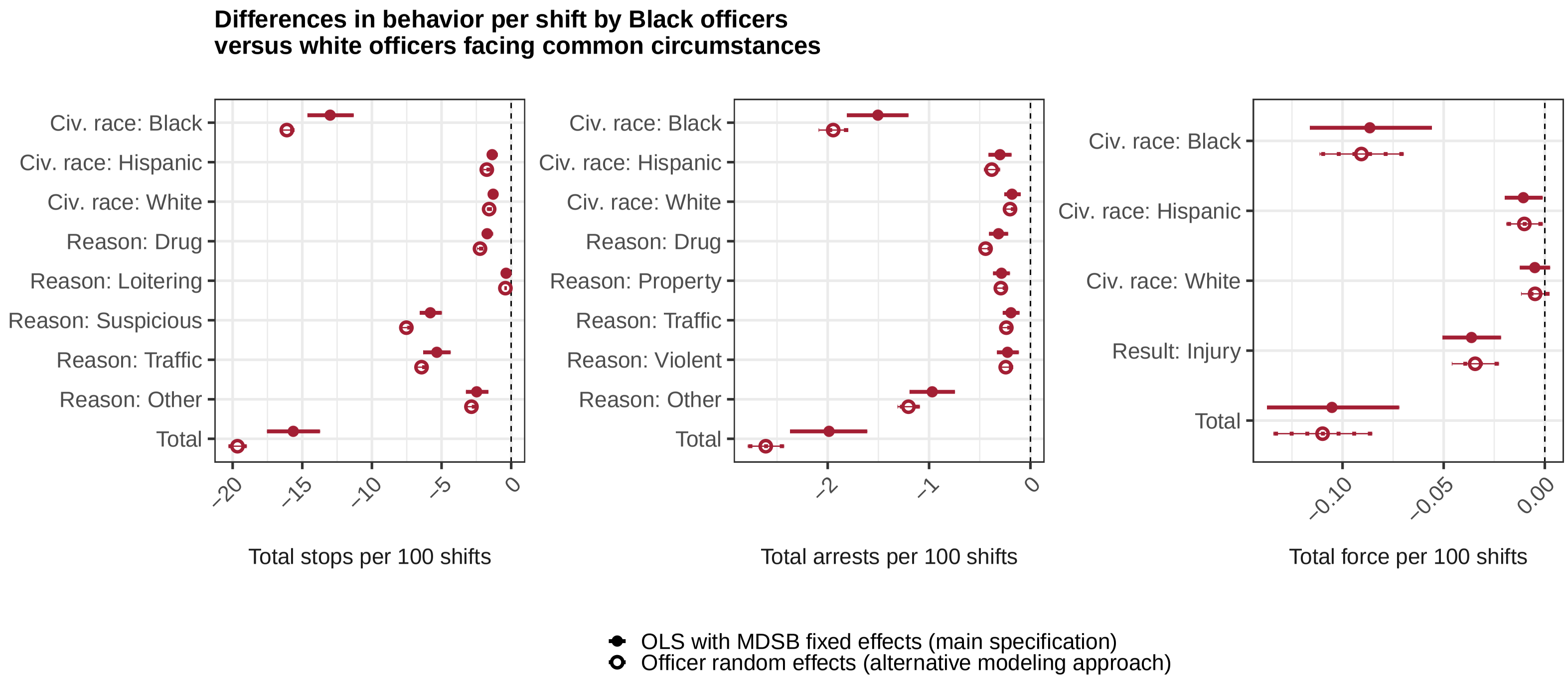

Figure S18. Comparison to random-effects estimator. Points (error bars) depict estimated differences (confidence intervals) in the number of uses of force by Black officers per shift, relative to White officers patrolling in the same month, day group (weekday/weekend), shift time (first/second/third watch), and beat. Results presented using two different approaches are presented: our primary specification with MDSB fixed effects and block-bootstrap confidence intervals, and an alternative specification using officer random effects.

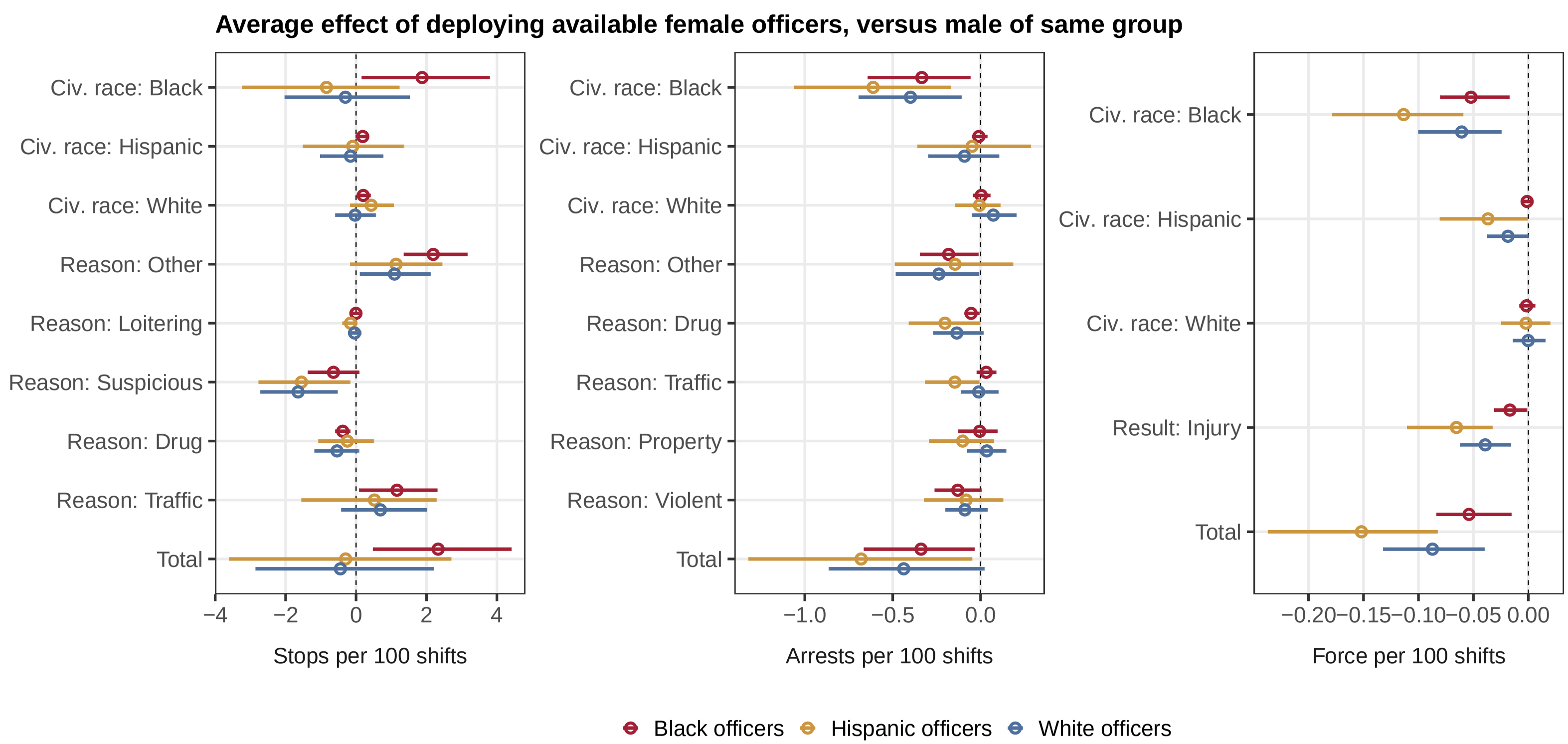

Figure S19. Within each racial/ethnic group, female and male officers behave differently when facing common circumstances. Each point and confidence interval reports a gender deployment effect for officers of a particular racial/ethnic group.

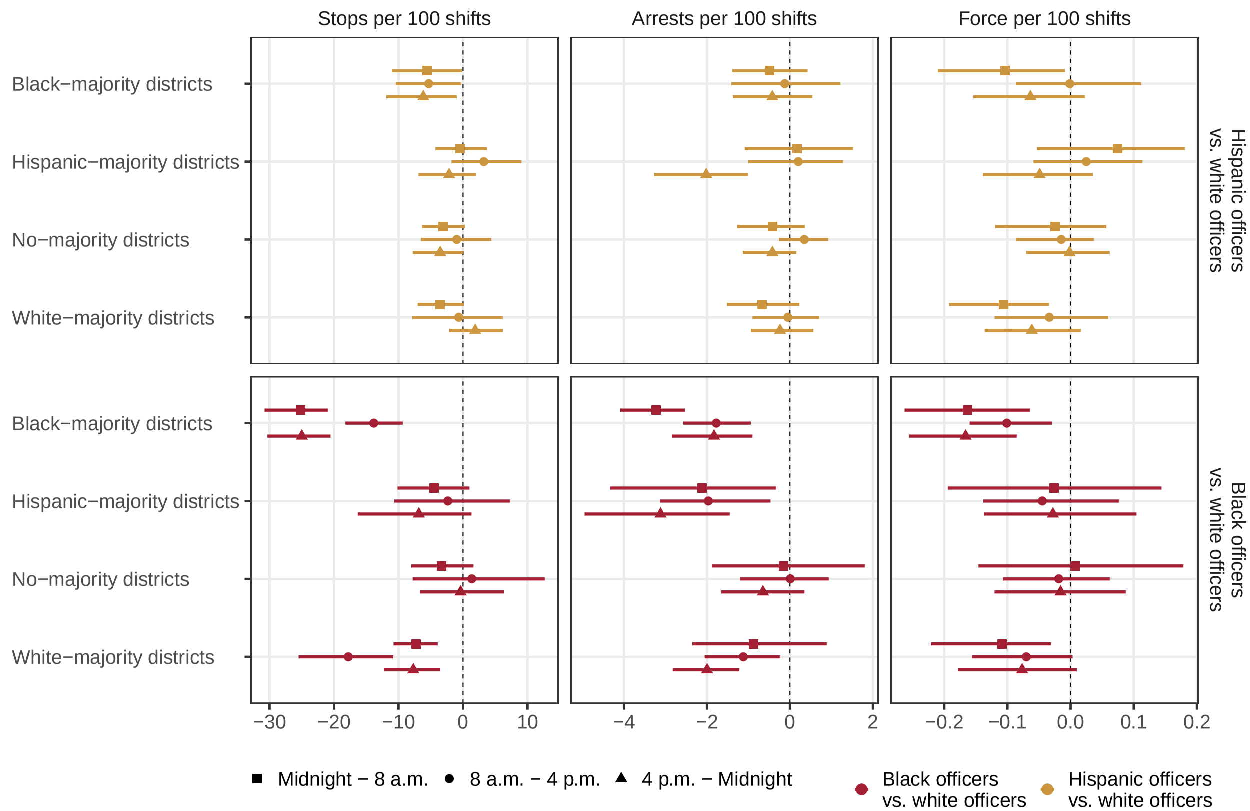

Figure S20. Racial deployment effects vary with district demographics and shift time. Each point and confidence interval reports a racial deployment effect for patrol assignments in a particular type of district (Black-, Hispanic-, White-, and no-majority) and shift time (starting at midnight, 8 a.m., and 4 p.m.).

Table S1. Average events per shift, by officer racial/ethnic group. Mean number of stops, arrests, and uses of force without adjustment for time or location. Typical behavior is reported for Black, Hispanic, and White officers individually, as well as the average pooling three officer races. Records associated with Native American/Alaskan and Asian/Pacific Islander officers are excluded due to small sample sizes. Officer behavior toward Native American/Alaskan and Asian/Pacific Islander civilians is not included for the purposes of computing total and reason-specific events. Values are scaled for ease of interpretation.

Behavior	Mean (Pooled)	Mean (Black off.)	Mean (Hisp. off.)	Mean (White off.)
Stops per 100 shifts:				
Civ. race: Black	30.27	26.07	30.86	32.45
Civ. race: Hispanic	9.23	2.11	14.22	10.97
Civ. race: White	5.66	2.10	6.13	7.53
Reason: Loitering	0.63	0.32	0.86	0.71
Reason: Suspicious	15.40	8.85	17.31	18.32
Reason: Drug	4.77	1.77	6.25	5.81
Reason: Traffic	11.80	8.06	13.91	12.97
Reason: Other	13.12	11.60	13.49	13.83
Total	45.72	30.59	51.81	51.64
Arrests per 100 shifts:				
Civ. race: Black	5.64	4.94	5.93	5.90
Civ. race: Hispanic	1.76	0.40	2.52	2.19
Civ. race: White	0.89	0.32	0.98	1.17
Reason: Drug	0.94	0.36	1.24	1.13
Reason: Traffic	0.70	0.35	0.78	0.87
Reason: Property	1.52	1.15	1.66	1.67
Reason: Violent	2.11	1.81	2.38	2.16
Reason: Other	3.09	2.04	3.46	3.53
Total	8.37	5.71	9.52	9.36
Force per 100 shifts:				
Civ. race: Black	0.20	0.14	0.21	0.23
Civ. race: Hispanic	0.04	0.01	0.06	0.05
Civ. race: White	0.03	0.01	0.03	0.04
Result: Injury	0.07	0.04	0.08	0.09
Total	0.27	0.16	0.30	0.32

Table S2. Average events per shift, by officer gender. Mean number of stops, arrests, and uses of force without adjustment for time or location. Typical behavior is reported for female and male officers separately, as well as the pooled average. Records associated with Native American/Alaskan and Asian/Pacific Islander officers are excluded for consistency with racial/ethnic analyses. Officer behavior toward Native American/Alaskan and Asian/Pacific Islander civilians is not included for the purposes of computing total and reason-specific events. Values are scaled for ease of interpretation.

Behavior	Mean (Pooled)	Mean (Female off.)	Mean (Male off.)
Stops per 100 shifts:			
Civ. race: Black	30.27	23.82	32.45
Civ. race: Hispanic	9.23	6.49	10.15
Civ. race: White	5.66	4.93	5.91
Reason: Loitering	0.63	0.42	0.71
Reason: Suspicious	15.40	10.56	17.04
Reason: Drug	4.77	2.66	5.48
Reason: Traffic	11.80	9.88	12.45
Reason: Other	13.12	12.28	13.40
Total	45.72	35.79	49.07
Arrests per 100 shifts:			
Civ. race: Black	5.64	4.03	6.18
Civ. race: Hispanic	1.76	1.11	1.98
Civ. race: White	0.89	0.69	0.95
Reason: Drug	0.94	0.47	1.09
Reason: Traffic	0.70	0.45	0.79
Reason: Property	1.52	1.21	1.63
Reason: Violent	2.11	1.68	2.26
Reason: Other	3.09	2.08	3.44
Total	8.37	5.90	9.20
Force per 100 shifts:			
Civ. race: Black	0.20	0.11	0.23
Civ. race: Hispanic	0.04	0.02	0.05
Civ. race: White	0.03	0.02	0.03
Result: Injury	0.07	0.03	0.09
Total	0.27	0.15	0.31

Table S3. Average events per shift for Hispanic officers, by language ability. Mean number of stops, arrests, and uses of force without adjustment for time or location. Typical behavior is reported for Spanish-speaking and non-Spanish-speaking Hispanic officers separately. Officer behavior toward Native American/Alaskan and Asian/Pacific Islander civilians is not included for the purposes of computing total and reason-specific events. Values are scaled for ease of interpretation.

Behavior	Mean (Non-Spanish Hisp. off.)	Mean (Spanish Hisp. off.)
Stops per 100 shifts:		
Civ. race: Black	39.10	21.41
Civ. race: Hispanic	13.55	14.99
Civ. race: White	5.81	6.49
Reason: Loitering	1.07	0.61
Reason: Suspicious	19.64	14.64
Reason: Drug	7.11	5.27
Reason: Traffic	16.89	10.49
Reason: Other	14.35	12.50
Total	59.06	43.50
Arrests per 100 shifts:		
Civ. race: Black	7.63	3.99
Civ. race: Hispanic	2.75	2.26
Civ. race: White	0.97	0.98
Reason: Drug	1.69	0.72
Reason: Traffic	1.06	0.46
Reason: Property	1.76	1.55
Reason: Violent	2.65	2.06
Reason: Other	4.27	2.53
Total	11.44	7.32
Force per 100 shifts:		
Civ. race: Black	0.28	0.14
Civ. race: Hispanic	0.06	0.05
Civ. race: White	0.03	0.03
Result: Injury	0.11	0.06
Total	0.37	0.22

Table S4. Effect of deploying Black officers versus White officers. The table displays average within-MDSB differences between various groups of officers relative to specified counterparts, as well as *p*-values adjusted for multiple testing. To roughly gauge the magnitude of effects, the far right column displays the result of dividing each effect by the volume of each enforcement activity exhibited by the reference group, citywide.

Outcome	Estimated effect	$p_{\text{adj.}}$	Typical ref.-group volume	Ratio
Stops per 100 shifts				
Civ. race: Black	−12.55	0.000	32.45	−0.39
Civ. race: Hispanic	−1.32	0.000	10.97	−0.12
Civ. race: White	−1.31	0.000	7.53	−0.17
Reason: Drug	−1.67	0.000	5.81	−0.29
Reason: Loitering	−0.34	0.000	0.71	−0.48
Reason: Suspicious	−5.72	0.000	18.32	−0.31
Reason: Traffic	−4.99	0.000	12.97	−0.38
Reason: Other	−2.44	0.000	13.83	−0.18
Total	−15.16	0.000	51.64	−0.29
Arrests per 100 shifts				
Civ. race: Black	−1.46	0.000	5.90	−0.25
Civ. race: Hispanic	−0.29	0.000	2.19	−0.13
Civ. race: White	−0.18	0.001	1.17	−0.16
Reason: Drug	−0.31	0.000	1.13	−0.27
Reason: Property	−0.29	0.000	1.67	−0.17
Reason: Traffic	−0.18	0.000	0.87	−0.21
Reason: Violent	−0.23	0.000	2.16	−0.10
Reason: Other	−0.93	0.000	3.53	−0.26
Total	−1.93	0.000	9.36	−0.21
Force per 100 shifts				
Civ. race: Black	−0.08	0.000	0.23	−0.38
Civ. race: Hispanic	−0.01	0.047	0.05	−0.20
Civ. race: White	−0.00	0.267	0.04	−0.13
Result: Injury	−0.03	0.000	0.09	−0.39
Total	−0.10	0.000	0.32	−0.32

Table S5. Effect of deploying Hispanic officers versus White officers. The table displays average within-MDSB differences between various groups of officers relative to specified counterparts, as well as p-values adjusted for multiple testing. To roughly gauge the magnitude of effects, the far right column displays the result of dividing each effect by the volume of each enforcement activity exhibited by the reference group, citywide.

Outcome	Estimated effect	$p_{\text{adj.}}$	Typical ref.-group volume	Ratio
Stops per 100 shifts				
Civ. race: Black	−2.90	0.000	32.45	−0.09
Civ. race: Hispanic	0.46	0.143	10.97	0.04
Civ. race: White	−0.43	0.010	7.53	−0.06
Reason: Drug	−0.53	0.007	5.81	−0.09
Reason: Loitering	0.05	0.500	0.71	0.07
Reason: Suspicious	−1.66	0.000	18.32	−0.09
Reason: Traffic	−0.46	0.283	12.97	−0.04
Reason: Other	−0.25	0.537	13.83	−0.02
Total	−2.84	0.001	51.64	−0.06
Arrests per 100 shifts				
Civ. race: Black	−0.30	0.022	5.90	−0.05
Civ. race: Hispanic	−0.07	0.292	2.19	−0.03
Civ. race: White	−0.08	0.018	1.17	−0.07
Reason: Drug	0.03	0.621	1.13	0.03
Reason: Property	−0.08	0.025	1.67	−0.05
Reason: Traffic	−0.08	0.041	0.87	−0.10
Reason: Violent	0.02	0.709	2.16	0.01
Reason: Other	−0.32	0.000	3.53	−0.09
Total	−0.44	0.012	9.36	−0.05
Force per 100 shifts				
Civ. race: Black	−0.03	0.013	0.23	−0.15
Civ. race: Hispanic	0.00	0.729	0.05	0.06
Civ. race: White	−0.01	0.232	0.04	−0.17
Result: Injury	−0.01	0.483	0.09	−0.06
Total	−0.04	0.021	0.32	−0.12

Table S6. Effect of deploying female officers versus male officers. The table displays average within-MDSB differences between various groups of officers relative to specified counterparts, as well as *p*-values adjusted for multiple testing. To roughly gauge the magnitude of effects, the far right column displays the result of dividing each effect by the volume of each enforcement activity exhibited by the reference group, citywide.

Outcome	Estimated effect	$p_{\text{adj.}}$	Typical ref.-group volume	Ratio
Stops per 100 shifts				
Civ. race: Black	−0.33	0.607	32.45	−0.01
Civ. race: Hispanic	−0.07	0.832	10.15	−0.01
Civ. race: White	0.22	0.176	5.91	0.04
Reason: Drug	−0.66	0.000	5.48	−0.12
Reason: Loitering	−0.06	0.195	0.71	−0.08
Reason: Suspicious	−1.61	0.000	17.04	−0.09
Reason: Traffic	0.66	0.126	12.45	0.05
Reason: Other	1.57	0.000	13.40	0.12
Total	−0.11	0.904	49.07	−0.00
Arrests per 100 shifts				
Civ. race: Black	−0.54	0.000	6.18	−0.09
Civ. race: Hispanic	−0.08	0.134	1.98	−0.04
Civ. race: White	0.02	0.538	0.95	0.02
Reason: Drug	−0.17	0.000	1.09	−0.15
Reason: Property	−0.02	0.487	1.63	−0.01
Reason: Traffic	−0.02	0.483	0.79	−0.03
Reason: Violent	−0.12	0.003	2.26	−0.05
Reason: Other	−0.27	0.000	3.44	−0.08
Total	−0.61	0.000	9.20	−0.07
Force per 100 shifts				
Civ. race: Black	−0.07	0.000	0.23	−0.31
Civ. race: Hispanic	−0.01	0.001	0.05	−0.28
Civ. race: White	−0.00	0.916	0.03	−0.01
Result: Injury	−0.03	0.000	0.09	−0.38
Total	−0.09	0.000	0.31	−0.28

Table S7. Effect of deploying Spanish-speaking versus non-speaking Hispanic officers officers. The table displays average within-MDSB differences between various groups of officers relative to specified counterparts, as well as p-values adjusted for multiple testing. To roughly gauge the magnitude of effects, the far right column displays the result of dividing each effect by the volume of each enforcement activity exhibited by the reference group, citywide.

Outcome	Estimated effect	$p_{\text{adj.}}$	Typical ref.-group volume	Ratio
Stops per 100 shifts				
Civ. race: Black	0.49	0.729	39.10	0.01
Civ. race: Hispanic	0.21	0.848	13.55	0.02
Civ. race: White	−0.03	0.848	5.81	−0.00
Reason: Drug	−0.23	0.705	7.11	−0.03
Reason: Loitering	−0.20	0.159	1.07	−0.19
Reason: Suspicious	−0.48	0.518	19.64	−0.02
Reason: Traffic	0.89	0.442	16.89	0.05
Reason: Other	0.74	0.278	14.35	0.05
Total	0.72	0.729	59.06	0.01
Arrests per 100 shifts				
Civ. race: Black	−0.42	0.046	7.63	−0.05
Civ. race: Hispanic	−0.20	0.278	2.75	−0.07
Civ. race: White	−0.02	0.746	0.97	−0.02
Reason: Drug	−0.13	0.232	1.69	−0.08
Reason: Property	−0.16	0.049	1.76	−0.09
Reason: Traffic	−0.12	0.130	1.06	−0.11
Reason: Violent	−0.10	0.383	2.65	−0.04
Reason: Other	−0.14	0.483	4.27	−0.03
Total	−0.66	0.037	11.44	−0.06
Force per 100 shifts				
Civ. race: Black	0.00	0.857	0.28	0.01
Civ. race: Hispanic	0.02	0.483	0.06	0.26
Civ. race: White	0.02	0.053	0.03	0.77
Result: Injury	0.02	0.371	0.11	0.20
Total	0.04	0.371	0.37	0.10

Table S8. Officers of differing demographic profiles behave differently when facing common circumstances. Each row reports a comparison between officers of various demographic groups. Each column contains average per-shift differences in a police behavior among officers assigned to the same beat and shift time, during the same month and day of week. Asterisks indicate Benjamini-Hochberg p-values accounting for officer-level clustering and adjusted for multiple testing of 92 hypotheses: * is < 0.05, ** is < 0.01, and *** is < 0.001. Alternative estimators, including adjusting for officer experience, are reported in SI Figures S15–S17 and produce substantively identical results.

Stops per 100 shifts

	Civilian race			Reason					
	Black	Hispanic	White	Drug	Loitering	Suspicious	Traffic	Other	Total
Black officers vs. White officers	-12.55***	-1.32***	-1.31***	-1.67***	-0.34***	-5.72***	-4.99***	-2.44***	-15.16***
Hispanic officers vs. White officers	-2.90***	0.46	-0.43*	-0.53**	0.05	-1.66***	-0.46	-0.25	-2.84**
Female officers vs. male officers	-0.33	-0.07	0.22	-0.66***	-0.06	-1.61***	0.66	1.57***	-0.11
Spanish-speaking Hispanic officers vs non-Spanish-speaking	0.49	0.21	-0.03	-0.23	-0.20	-0.48	0.89	0.74	0.72

Arrests per 100 shifts

	Civilian race:			Reason:					
	Black	Hispanic	White	Drug	Property	Traffic	Violent	Other	Total
Black officers vs. White officers	-1.46***	-0.29***	-0.18**	-0.31***	-0.29***	-0.18***	-0.23***	-0.93***	-1.93***
Hispanic officers vs. White officers	-0.30*	-0.07	-0.08*	0.03	-0.08*	-0.08*	0.02	-0.32***	-0.44*
Female officers vs. male officers	-0.54***	-0.08	0.02	-0.17***	-0.02	-0.02	-0.12**	-0.27***	-0.61***
Spanish-speaking Hispanic officers vs non-Spanish-speaking	-0.42*	-0.20	-0.02	-0.13	-0.16*	-0.12	-0.10	-0.14	-0.66*

Uses of force per 100 shifts

	Civilian race			Result	
	Black	Hispanic	White	Injury	Total
Black officers vs. White officers	-0.08***	-0.01*	-0.00	-0.03***	-0.10***
Hispanic officers vs. White officers	-0.03*	0.00	-0.01	-0.01	-0.04*
Female officers vs. male officers	-0.07***	-0.01**	-0.00	-0.03***	-0.09***
Spanish-speaking Hispanic officers vs non-Spanish-speaking	0.00	0.02	0.02	0.02	0.04

References and Notes

1. M. Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (The New Press, 2010).

2. A. Lerman, V. Weaver, *Arresting Citizenship: The Democratic Consequences of American Crime Control* (Univ. of Chicago Press, 2014).

3. J. Forman Jr., *Locking Up Our Own: Crime and Punishment in Black America* (Farrar, Straus and Giroux, 2017).

4. U.S. Department of Justice, "Advancing Diversity in Law Enforcement" (2016).

5. A. Fantz, C. Tolan, "Want to reform the police? Hire more women," CNN*,* 2020.

6. Z. Qian, Race and social distance: Intermarriage with non-Latino Whites. *Race Soc.* **5**, 33–47 (2002). doi:10.1016/j.racsoc.2003.12.003

7. T. A. LaVeist, A. Nuru-Jeter, Is doctor-patient race concordance associated with greater satisfaction with care? *J. Health Soc. Behav.* **43**, 296–306 (2002). doi:10.2307/3090205 Medline

8. E. M. Kearns, Why are some officers more supportive of community policing with minorities than others? *Justice Q.* **34**, 1213–1245 (2017). doi:10.1080/07418825.2017.1380837

9. H. Tajfel, J. Turner, ''An Integrative Theory of Intergroup Conflict,'' in *The Social Psychology of Intergroup Relations*, W. G. Austin, S. Worchel, Eds. (Wadsworth, 1979), pp. 55–65.

10. C. Neil Macrae, G. V. Bodenhausen, Social cognition: Thinking categorically about others. *Annu. Rev. Psychol.* **51**, 93-120 (2000).

11. S. T.Fiske, What we know now about bias and intergroup conflict, the problem of the century. *Curr. Directions Psychol. Sci.* **11**, 123–128 (2002).

12. J. D. Clinton, D. E. Lewis, Expert opinion, agency characteristics, and agency preferences. *Polit. Anal.* **16**, 3–20 (2008). doi:10.1093/pan/mpm009

13. D. E. Lewis, Measurement and public service motivation: New insights, old questions. *Int. Public Manage. J.* **13**, 46–55 (2010). doi:10.1080/10967490903547233

14. E. Linos, N. Riesch, Thick Red Tape and the Thin Blue Line: A Field Study on Reducing Administrative Burden in Police Recruitment. *Public Admin. Rev* **80**, 92–103 (2020).

15. J. Q. Wilson, *Bureaucracy: What Government Agencies Do and Why They Do It*, J. Q. Wilson, Ed. (Basic Books, 1989).

16. K. J. Meier, Representative bureaucracy: An empirical analysis. *Am. Polit. Sci. Rev.* **69**, 526–542 (1975). doi:10.2307/1959084

17. V. Ray, A Theory of Racialized Organizations. *Am. Sociol. Rev.* **84**, 26–53 (2019). doi:10.1177/0003122418822335

18. S. H. Alim, G. Smitherman, *Articulate While Black: Barack Obama, Language, and Race in the U.S.* (Oxford Univ. Press, 2012).

19. J. Legewie, J. Fagan, “Group threat, police officer diversity and the deadly use of police force,” *Columbia Public Law Research Paper* (2016).

20. J. J. Donohue III, S. D. Levitt, The impact of race on policing and arrests. *J. Law Econ.* **44**, 367–394 (2001). doi:10.1086/322810

21. B. Smith, The Impact of Police Officer Diversity on Police-Caused Homicides. *Policy Stud.* **31**, 147–162 (2003). doi:10.1111/1541-0072.t01-1-00009

22. A. Fagan, A. Braga, R. K. Brunson, A. Pattavina, Stops and stares: Street stops, surveillance, and race in the new policing. *Fordham Urban Law J.* **43**, 539–614 (2016).

23. P. B. Hoffman, E. R. Hickey, Use of force by female police officers. *J. Crim. Justice* **33**, 145–151 (2005). doi:10.1016/j.jcrimjus.2004.12.006

24. E. K. Weisburst, Police use of force as an extension of arrests: Examining disparities across civilian and officer race. *AEA Papers and Proceedings* **109**, 152–156 (2019). doi:10.1257/pandp.20191028

25. J. West, “Racial Bias in Police Investigations,” Working Paper, 2018; https://people.ucsc.edu/~jwest1/articles/West_RacialBiasPolice.pdf.

26. J. McCrary, The effect of court-ordered hiring quotas on the composition and quality of police. *Am. Econ. Rev.* **97**, 318–353 (2007). doi:10.1257/aer.97.1.318

27. A. R. Miller, C. Segal, Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence. *Rev. Econ. Stud.* **86**, 2220–2247 (2019). doi:10.1093/restud/rdy051

28. A. Harvey, T. Mattia, “Reducing Racial Disparities in Crime Victimization,” Working Paper, 2019; https://wp.nyu.edu/annaharvey/wp-content/uploads/sites/6417/2020/07/Victimization_Revised.pdf.

29. M. Garner, A. Harvey, H. Johnson, Estimating Effects of Affirmative Action in Policing: A Replication and Extension. *Int. Rev. Law Econ.* **62**, 105881 (2020). doi:10.1016/j.irle.2019.105881

30. D. A. Sklansky, Not your father’s police department: Making sense of the new demographics of law enforcement. *J. Crim. Law Criminol.* **96**, 1225 (2005).

31. ACLU, “Stop and Frisk in Chicago,” American Civil Liberties Union of Illinois, 2015.

32. J. Byrne, R. Pearson, J. Gorner, “Mayor Rahm Emanuel rejects President Donald Trump’s call for Chicago police to use stop-and-frisk tactics,” *Chicago Tribune*, 9 October 2018.

33. Police Accountability Task Force, “Recommendations for Reform: Restoring Trust Between the Chicago Police and the Communities They Serve,” 2016.

34. E. McClelland, “Chicago’s Top Cop Was Fired After the Laquan McDonald Shooting. Now He Wants to Be Mayor,”*Politico*, 29 September 2018.

35. WTTW, “Power, Politics, & Pride: Afro-American Patrolmen's League,” https://interactive.wttw.com/dusable-to-obama/afro-american-patrolmens-league [accessed 28 January 2021].

36. M. M. Adams, “The Patrolmen’s Revolt: Chicago Police and the Labor and Urban Crises of the Late Twentieth Century,” thesis, University of California, Berkeley (2012).

37. B. A. Ba, “Going the Extra Mile: the Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago,” Working paper, 2019.

38. N. Rim, R. G. Rivera, B. A. Ba, In-group Bias and the Police: Evidence from Award Nominations (2019). Faculty Scholarship at Penn Law. 2132. https://scholarship.law.upenn.edu/faculty_scholarship/2132.

39. P. Atiba Goff, K. Barsamian Kahn, Racial bias in policing: Why we know less than we should. *Soc. Issues Policy Rev.* **6**, 177–210 (2012). doi:10.1111/j.1751-2409.2011.01039.x

40. B. Ba, D. Knox, J. Mummolo, R. Rivera, The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago [Source Code]. (2021). https://doi.org/10.24433/CO.4519202.v1.

41. D. Holliday, “Logan Square CAPS Meeting Set for Thursday Night,” *DNAinfo,* 2015.

42. C. P. Dept, “Annual Watch and Furlough Selections, Vacation Schedules, and 2020 Operations Calendar” (2019); http://directives.chicagopolice.org/directives/data/a7a57b85-16df962c-bf316-df96-5e9d14d38fd3880c.html?hl=true.

43. C. P. Dept, “2015 Operations Calendar”; https://web.archive.org/web/20151101000000*/http://www.chicagofop.org/operations-calendars [accessed 5 October 2020].

44. C. P. Dept, “Annual Report” (2008); https://home.chicagopolice.org/wp-content/uploads/2014/12/2008-Annual-Report.pdf.

45. D. B. Rubin, Formal mode of statistical inference for causal effects. *J. Stat. Plan. Inference* **25**, 279–292 (1990). doi:10.1016/0378-3758(90)90077-8

46. D. B. Rubin, Estimating causal effects of treatments in randomized and non-randomized studies. *J. Educ. Psychol.* **66**, 688–701 (1974). doi:10.1037/h0037350