Diversity in Policing: The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago[*]

Bocar Ba[†] Dean Knox[‡] Jonathan Mummolo[§] Roman Rivera[¶]

July 1, 2020

[*]Data for this project were gathered in collaboration with Invisible Institute, with the help of Sam Stecklow and Emma Herman. We thank Rachel Ryley and the Lucy Parsons Lab.

[†]Assistant Professor of Economics, University of California Irvine, bba@uci.edu.

[‡]Assistant Professor of Operations, Information and Decisions, Wharton School, University of Pennsylvania, dcknox@upenn.edu.

[§]Assistant Professor of Politics and Public Affairs, Princeton University, jmummolo@princeton.edu.

[¶]PhD student, Dept. of Economics at Columbia University, r.g.rivera@columbia.edu.

Abstract

A history of exclusively white male police forces in the U.S. has made diversifying personnel one of the oldest and most often proposed police reforms, but data challenges have precluded micro-level evaluations of its impact. Using newly collected personnel data and millions of ultra-fine-grained records on officer deployment and behavior, we conduct a detailed quantitative case study of diversity in the Chicago Police Department. We show how officers from marginalized groups are consistently assigned to different working conditions than white and male officers, meaning they typically encounter vastly different circumstances and civilian behaviors. As a result, coarse agency- or district-level analyses often fall short of all-else-equal comparisons between officer groups, making it difficult to draw firm conclusions about the effects of diversification policies. To assess the effects of deploying officers with various racial/ethnic and gender profiles, we leverage detailed records of daily patrol assignments to evaluate officers against counterparts working in the same collections of city blocks, in the same month and day of week, and at the same time of day. Compared with white officers facing identical conditions, we show Black and Hispanic officers both make substantially fewer stops, arrests, and use force less often, especially against Black civilians. Much of the gaps in stops and arrests are due to a decreased focus on discretionary contact, such as stops for vaguely defined "suspicious behavior." Hispanic and white officers exhibit highly similar behavior toward Hispanic civilians, though we show Hispanic officers who speak Spanish make fewer arrests in general than those who do not speak Spanish. Within all racial/ethnic groups, female officers are substantially less likely to use force relative to male officers. Taken together, these results show the substantial impact of diversity on police treatment of minority communities, and emphasize the need to consider multiple facets of police officers when crafting personnel-driven reforms.

Racial disparities in police-civilian interactions and high-profile incidents of excessive force continue to fuel allegations of abusive and discriminatory policing [2, 25]. Central to these critiques are the fact that throughout the history of policing in the U.S.—from the formation of the first organized security patrols on slave plantations [33, 42], to the recent concentrated implementation of aggressive tactics like "Stop, Question and Frisk" in communities of color [14]—many police forces in the U.S. have been nearly all white and male [11]. In turn, some of the most frequently proposed reforms aimed at reducing inequities and police brutality have centered on hiring more nonwhite [7] and female [10] officers. One agency that has undergone substantial diversification in recent decades is the Chicago Police Department (CPD), transforming from a nearly all white male agency to one in which half of sworn officers are minorities and over one-fifth are female. This history, combined with unprecedented data availability on the daily geographic assignment and behavior of officers, allows for a thorough assessment of the practical consequences of diversity in law enforcement. In this paper, we examine the Chicago case in depth to provide some of the most credible micro-level evidence to date on the impact of deploying officers of differing racial/ethnic and gender groups on police interactions with civilians.

A central obstacle to rigorous evaluation of personnel reforms in policing has been the lack of sufficiently fine-grained data on officer deployment and behavior. Most studies rely on inherently limited cross-jurisdiction analyses [e.g. 24, 8, 40], comparing aggregate outcomes like arrests and uses of force between agencies with differing levels of diversity. This approach often cannot distinguish whether apparent differences in officer behavior arise due to selection (e.g., agencies that choose to diversify may differ on unmeasured traits) or ecological fallacies (e.g., white and nonwhite officers within jurisdictions are assigned to work in different circumstances—which we demonstrate is in fact true). While strategies like agency-level pre-/post-reform comparisons have been employed to combat these sources of bias [27, 28, 12, 16], these aggregated analyses mask details of police-civilian encounters, precluding explorations of important sources of heterogeneity. In cases where researchers have been able to obtain micro-level data on officers, analyses are often limited in scope (e.g. traffic accidents [46] or 911 calls [19, 45]), rely on difficult-to-validate survey responses [32, 47], or otherwise lack necessary data to make reliable inferences. Specifically, many studies of officer race and gender that rely on administrative data examine records of enforcement activities *only* [e.g. logs of pedestrian stops or arrests; 3, 9, 20, 34, 44], omitting data from shifts in which officers take no such actions—introducing potentially severe selection bias [17, 18, 22]. Using available (but often limited) data, prior work has shown suggestive evidence that female officers are less likely to use force [e.g. 26, 31] though conclusions vary [20]. Findings with regard to racial diversity have been decidedly mixed: in an exhaustive review of the empirical literature on racial diversity in policing, one prominent legal scholar concluded: "[t]he fairest summary of the evidence is probably that we simply do not know" [39, 1225].

In this paper, we draw on a host of newly collected data sets that together allow us to overcome these longstanding limitations. Our data, which include personnel records covering roughly 30,000 Chicago police officers, was assembled through years of open records requests and lawsuits; they include officer demographics, language skills, shift assignments and career progression. We use these records to paint a detailed picture of the diversification of the department and how it has deployed officers from marginalized groups. We find substantial variation in deployment patterns across officers of different races and genders.[1] For example, Black officers are more often deployed in high-crime areas, and female officers tend to work during different times of day. These differential deployment patterns highlight a source of confounding present in many agency-level analyses of personnel reforms that finer-grained data can address.

We link these personnel files to timestamped, geolocated records of officers' decisions to stop, arrest, and use force against civilians. After aggressively pruning these data to maximize the validity of our analyses, we are left with a panel containing 2.9 million officer-shifts and 1.6 million detailed observations of enforcement events. Most importantly, we leverage fine-grained information on officers' daily patrol assignments—the specific times and beats (roughly, collections of city blocks) that they are assigned to patrol—to examine how officers of different groups behave when faced with identical circumstances and civilian behaviors. Crucially, this data sheds light not only on when officers take enforcement action against civilians, but also accounts for decisions *not* to take action—thus avoiding a major source of selection bias.

Our analysis does not claim to identify the causal "effect" of manipulating officers' identity; minority and female officers may differ from their counterparts in many unmeasured ways. Rather, we seek to estimate a far more policy-relevant quantity: the effect of *deploying* available officers of one group (with the totality of their traits), relative to another group, on police treatment of civilians. This quantity is much more feasibly estimated because, as we explain below, it only requires adjusting for external environmental factors like the civilian behaviors faced—not netting out internal officer traits. Our analysis therefore sheds light on the effect of a feasible and often proposed policy reform, and it demonstrates what behavior civilians can expect from police when officers of a given demographic profile are deployed, holding environmental factors constant.

Our results highlight the need for much greater collection of policing data, and they suggest deploying more women and officers of color (relative to white males) would reduce the amount of enforcement activity and physical force endured by civilians, especially Black civilians. At a high level, we find:

[1] It is unclear whether Chicago consistently differentiates between sex and gender in administrative data. We use the term "gender" throughout in keeping with officer personnel files we received in response to open records requests.

(1) Minority officers are assigned to very different districts, and even within districts, receive vastly different geographic and temporal patrol assignments. Without accounting for these differences in working conditions, there is no way to meaningfully characterize the differences in behavior between white/minority and male/female officers, suggesting a need for greater transparency by police departments to enable careful evaluation of officer behavior.

(2) Compared to white officers working in the same places and times, Black officers make significantly fewer stops and arrests, and they use force less often. These differences are substantial, equivalent to 31%, 22%, and 35% of typical white-officer activity, respectively. Examining a wide range of officer activity, we show this is mostly driven by a decrease in discretionary activity (e.g. 38% of the reduction in stops are due to a reduced focus on vaguely defined "suspicious behavior"), rather than lower enforcement of severe criminal behavior (decreased violent-crime arrests account for only 12% of the overall Black-white gap in arrest volume). Moreover, higher levels of enforcement by white officers fall primarily on marginalized groups: approximately 80% of the gap in officer stops, arrests, and uses of force is due to differing treatment of Black civilians. Hispanic officers display similar reductions in enforcement activity, relative to white officers, but to a lesser extent than their Black counterparts. In addition, we find Hispanic officers who speak Spanish make fewer arrests than their non-Spanish-speaking counterparts, underscoring an additional and previously under-explored, but potentially consequential, officer attribute.

(3) Relative to male officers facing identical circumstances, female officers make somewhat fewer arrests (differences equivalent to 7% of typical male arrest rates) and use force dramatically less (31%); moreover, over 80% of this gap is due to a reduced focus on Black civilians. This holds true even when comparing within racial/ethnic groups: across the board, Black, Hispanic, and white policewomen use force less often than their co-racial/co-ethnic male counterparts, with reductions that are concentrated nearly entirely among interactions with nonwhite civilians.

Overall, our results suggest that deploying officers of various races and genders yields substantial differences in police treatment of civilians. However, our findings also suggest the need to consider multiple facets of police officers when crafting personnel-driven reforms. Like the civilians they police, officers are complex, not easily reducible to single demographic traits. Scholars and reformers must take stock of the whole officer when considering the impact of diversity initiatives on police-civilian interactions.

All data and interactive replication materials are publicly available at [LINK]. We encourage readers to probe and extend our analyses.

1 The Case of the Chicago Police Department

Our analysis is confined to records from a single city, a feature that affords us unusually detailed data at the expense of geographic scope. Given the substantial advance in accurate assessment of diversification policies, we view the tradeoff as worthwhile. This analysis also provides a template for future data collection and research in other jurisdictions.

Chicago represents a valuable opportunity to evaluate the effects of deploying diverse officers. It is a large and racially diverse metropolis, with more than half the city's population identifying as nonwhite. The police force has also racially diversified in recent decades, with roughly 22% of officers identifying as Black, 23% Hispanic, and 3% Asian.[2] The agency's officers are currently 22% female, a stark change from its 99% male composition in 1970. The city is also heavily racially segregated,[3] has come under scrutiny in recent years for applying a range of controversial aggressive policing tactics such as "Stop and Frisk" on a wide scale, and has received national attention for the killing of Laquan McDonald and ensuing social unrest [6]. McDonald, 17, was shot and killed by Chicago police officer Jason Van Dyke. Following the delayed release of dash-cam footage of the killing, public outrage over the incident led to protests, a DOJ investigation, the election of a new Cook County State's Attorney, and Van Dyke being found guilty of second degree murder. The existence of these problems makes Chicago an important test case for proposed reforms, as it is arguably among the major jurisdictions in which reform is sorely needed.

1.1 Data on Agency Personnel

To understand the history of how the CPD diversified over time, we rely on both published accounts of department history and newly acquired quantitative data. Over a period of three years, we submitted open records requests to the CPD and the city's Department of Human Resources seeking data on officer demographics and behavior. The resulting records include the name, race, gender, birth year, salary, language skill, unit assignments and appointment date of each officer [5, 35]. These 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 demographics cover only 1995–2010 and 2016–2017.[4] Using these records, we extend the time-series backward to 1970, allowing for a comprehensive descriptive portrait of the evolution of the demographic correspondence between

[2] These figures describe the racial distribution of the CPD in 2016 according to our personnel records; see Figure 1 for additional details.

[3] See `http://www.censusscope.org/us/print_rank_dissimilarity_white_black.html`

[4] The CPD Annual Reports from 1966 to 1969 included counts of "police women" and "police matrons," but reporting on these statistics was discontinued after 1969.

CPD personnel and city residents.[5]

Figure 1 shows how the CPD, which currently employs about 12,000 sworn officers, slowly diversified over time. As [27] recounts, before a series of lawsuits, the CPD was slightly less than than 20% Black, in a city with a one-third Black population in 1970 [27]. In the early 1970s, the Afro-American Patrolmen's League (AAPL) filed a discrimination suit against the CPD "on hiring, promotion, assignment, and discipline" [30], with the DOJ soon joining them [27]. In 1974, hiring quotas were imposed, and Black hiring shares increased from 10% to 40% by 1975—though CPD composition changed more slowly due to low turnover [27]. These reforms also altered the gender composition of the department within racial groups and across ranks; women made up a larger proportion of Black recruits, initially leading white women to trail Black women in new hiring and promotions [1, 92].

While the CPD has made strides in moving its demographic profile toward parity with the city's, there exists substantial divergence in *working conditions*. Figure 2 displays the average characteristics of the districts—the 22 geographic regions delineated by the CPD—to which officer groups are assigned. While the figure shows only minimal gendered variation in district assignments, the differences associated with officer race/ethnicity are stark. Black officers are assigned to districts with roughly 50% higher rates of violent crime, and perhaps most strikingly, Black officers are assigned to districts with large co-racial resident populations—on average nearly 75% co-racial, far higher than the average 25–30% co-racial/co-ethnic districts where white and Hispanic officers serve. (SI Appendix B.1 contains a detailed discussion how districts are defined and analysis of additional district characteristics. SI Figure 2 presents average district characteristics disaggregated by year. SI Figure A2 shows that the demographic profile of assigned officers somewhat tracks that of district residents, but officers are disproportionately white.)

In supplemental analyses (SI Figures A4–A5) we also show considerable variation in assigned shift times. For example, 45% of standard shifts served by female officers are on third watch (4 p.m. to midnight), compared to 55% for male officers. Similarly, 43% of standard shifts served by Black officers are third watch, compared to 55% of white-officer shifts and 60% for Hispanic officers. Moreover, Figures A6–A7 demonstrate that marginalized groups are tasked with patrolling different sets of beats, compared to white or male officers serving in the same district. (All p-values < 0.001.)

These patterns underscore a central difficulty in testing whether officers with different demographic profiles perform their duties differently. Namely, white officers work in different environments from Black officers, on average (especially with regard to local racial composition), and men and women work during different hours of the day. This means any inferred differences in officer behavior that rely on data aggregated to large geographic units and time periods—such

[5] See SI Section A.2 for details on the coding on race/ethnicity.

Figure 1: **Composition of CPD officers and city residents over time.** Red, blue, and black ◦ respectively 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.

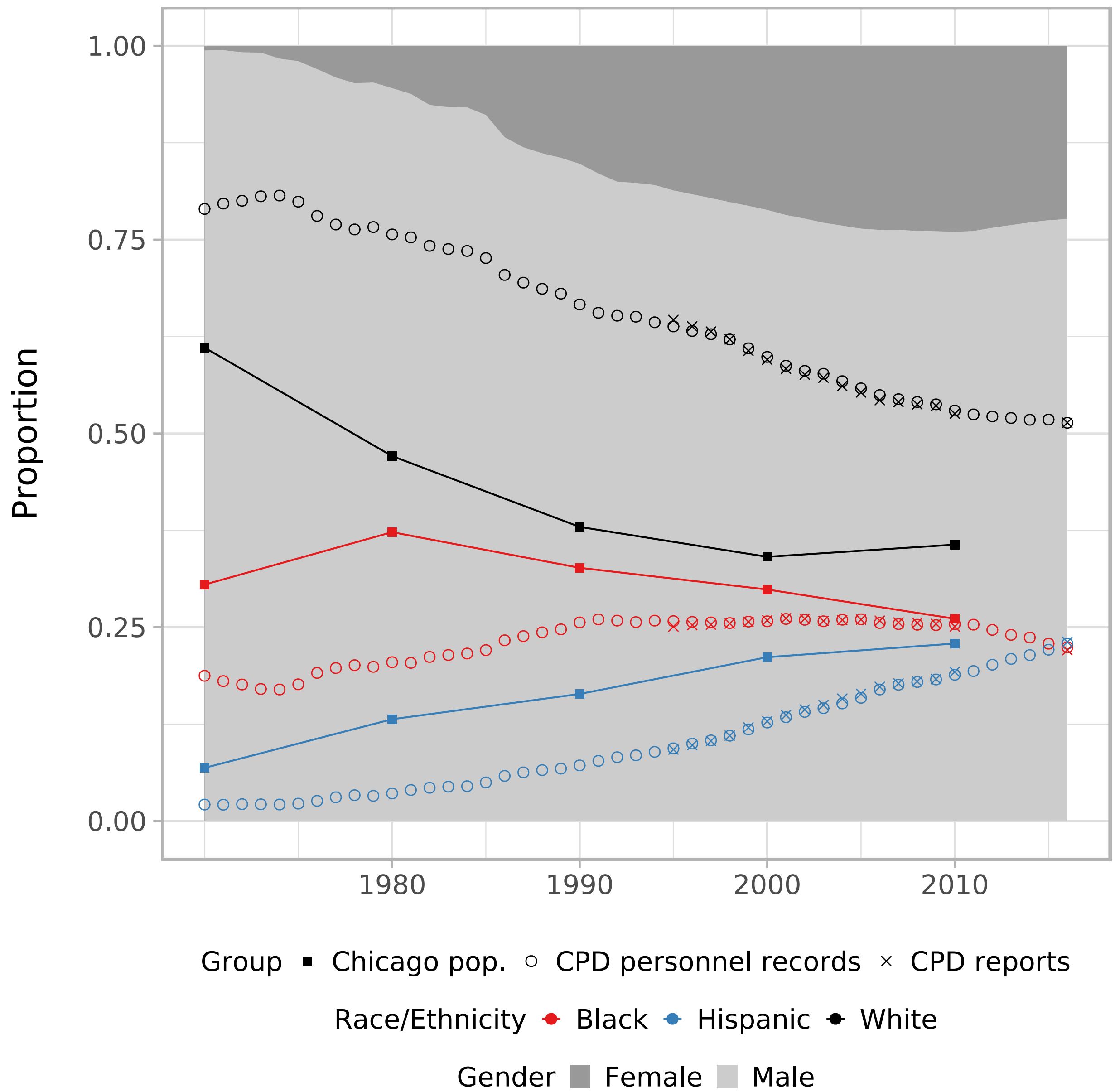
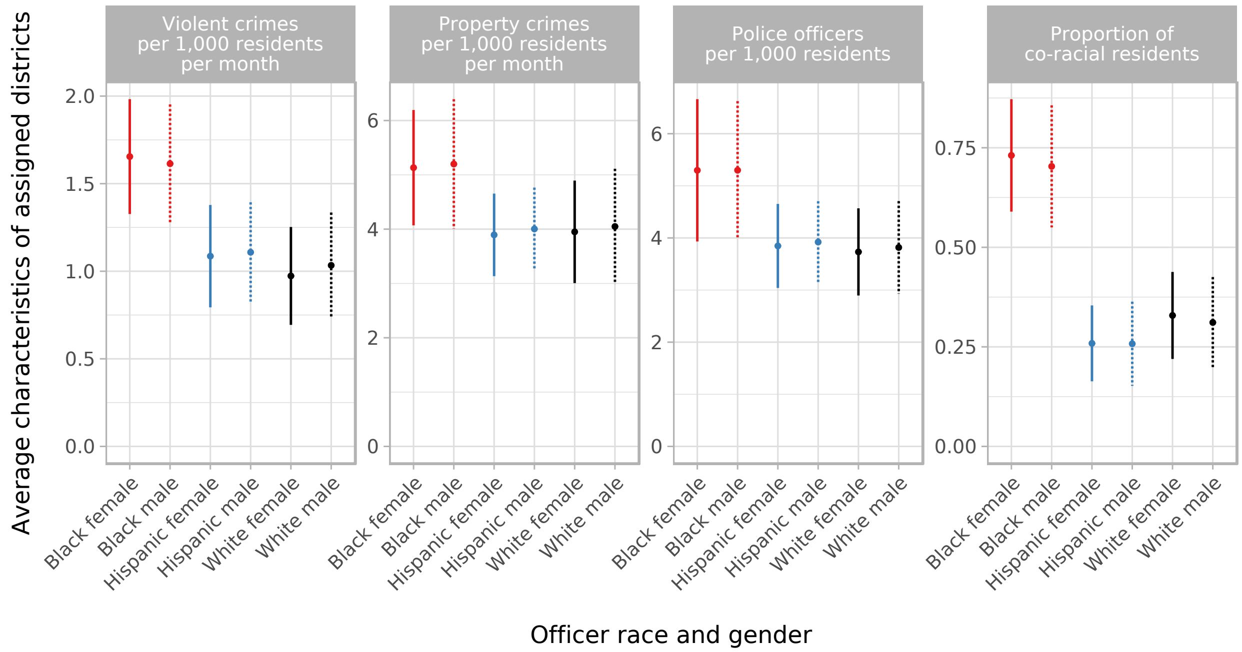


Figure 2: **Characteristics of assigned districts.** Average features of assigned geographic units are plotted for various officer groups, based on 1,089,707 monthly assignments from 2006–2016. Confidence intervals cluster on district-month.



as district-months or agency-years, the analytic strategy in much prior work—may simply reflect differing patrol environments rather than inherent differences in their approaches to policing. Fine-grained assignment data overcomes this central obstacle to inference.

2 Data on Officer Behavior

Our data also contain detailed data on officers' stops, arrests and uses of force against civilians. We merge these with records of officers' daily beat and shift assignments[6] and U.S. Census data. Together these data provide an unprecedentedly detailed view of the day-to-day behavior of officers in a major U.S. law enforcement agency over an extended period. Table 1 contains aggregate information on stops, arrests, and uses of force from Jan. 2012 to Dec. 2015 (the period covered by our behavioral analysis) by officer group. Due to the small sample sizes associated with groups including Asian Americans and Native Americans, our analysis is limited to Black, Hispanic, and white officers (together accounting for roughly 97% of officers for whom shift data is available).

Figure 3 illustrates the dataset's various attributes by highlighting a small slice of its temporal and geographic coverage. The figure maps activity during a four-month window in the CPD's

[6]We are grateful to Lucy Parsons Labs for publicly releasing data on civilian stops and Rachel Ryley for generously sharing data on beat assignments.

Wentworth District (District 2), a highly segregated 7.5 square mile territory on Chicago's South Side that is 96% Black and consistently ranks among the city's most violent districts in per-capita crime rates. The district is comprised of 15 patrol areas[7] which are shaded according to their racial composition. Points distinguish between geolocated stops, arrests, and uses of force during this period. The figure also describes the behavior of four CPD officers working in District 2 during this time, renamed to maintain anonymity. For example, "Officer A" is female, Hispanic, speaks both English and Spanish, and was born in 1965; "Officer D" is a white male born in 1981 who does not speak Spanish. The figure shows the specific beats to which officers were assigned over time and each officer's behavior while working in these beats.

[7]Based on the CPD 2008 Annual Report.

Figure 3: **Detailed view of the data.** The right panel figure maps police activity in a single police district (Wentworth, CPD 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 four-month period, with panel headers indicating age, gender, ethnicity/race, and language ability. Encircled incidents are described further in the left middle panel, which report 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.

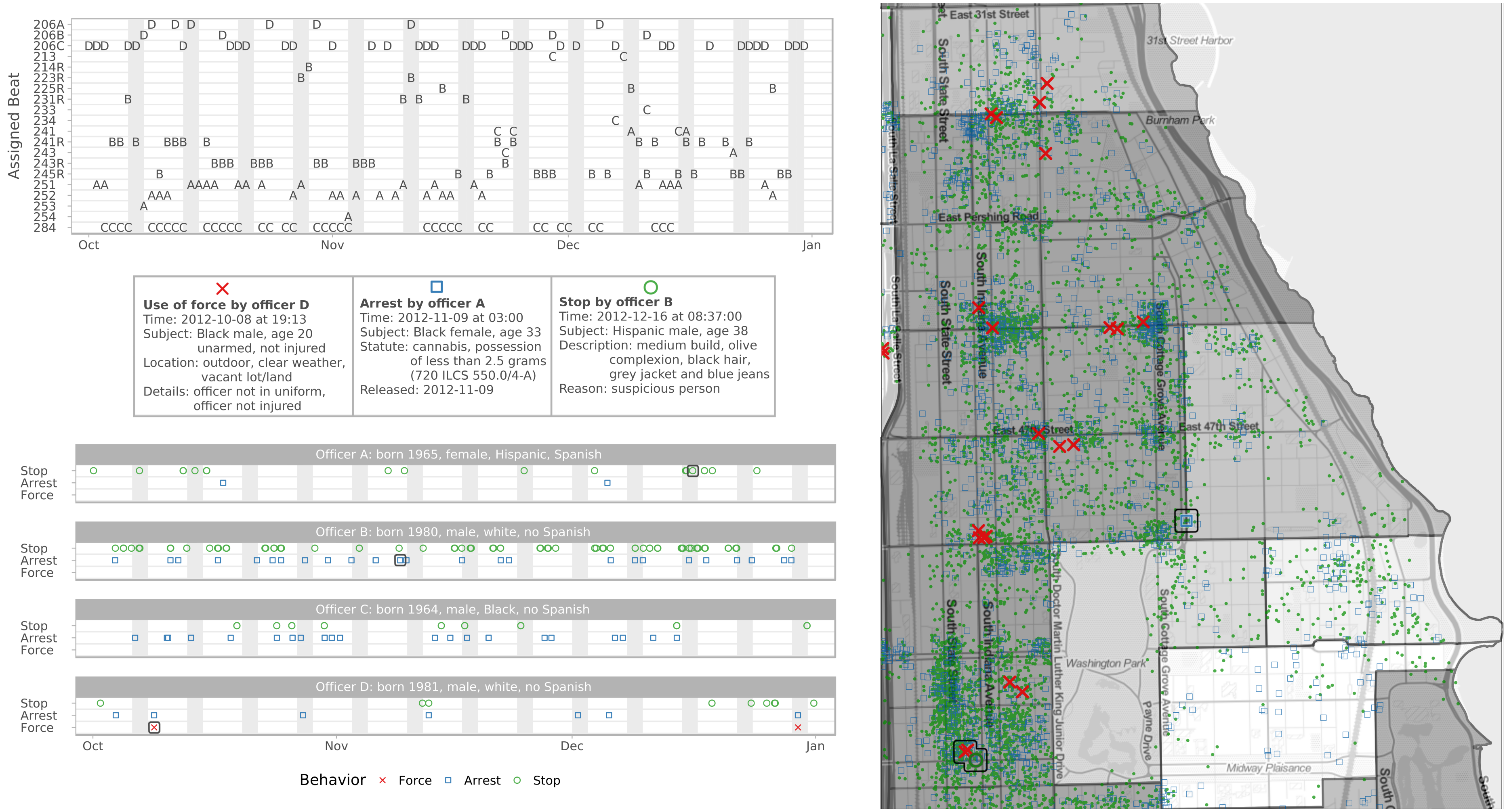


Table 1: **Summary of data on officer behavior (counts), 2012-2015.** Summary statistics are reported after pruning officers, shifts, and event records aggressively to ensure common circumstances in our behavioral analysis, as described in the main text.

	Black officers	Hispanic officers	White officers	Female officers	Male officers
Stops	253,609	356,541	729,078	264,552	1,074,676
Arrests	47,406	65,587	132,285	43,630	201,648
Uses of force	1,355	2,081	4,514	1,125	6,825
Shifts	830,062	689,239	1,413,977	740,240	2,193,038
Officers	1,835	1,674	3,440	1,786	5,163

3 Research Design: Assessing the Impact of Deployment

These granular data permit precise comparisons between officers facing identical working conditions. We assemble a panel data set in which each row represents an officer-shift—an assigned eight-hour patrol period—and contains detailed information on the officer's actions and their context.[8] In each of these 2.9 million patrol assignments, we compute the officer's profile of stops, arrests, and uses of force. Officers of different demographic profiles are then compared to peers working in the same unique combination of month, day of week, shift time, and beat (roughly a small collection of city blocks, averaging 0.82 square miles citywide)—a narrow slice of time and space that we abbreviate "MDSBs."

By making these precise comparisons, we ensure observed differences in officer behavior are not due to disparities in external conditions. That is, we can safely assume all officers in an MDSB face the same set of average enforcement opportunities, civilian activity, neighborhood attributes such as infrastructure and architecture, and time-varying conditions such as weather and lighting. Put differently, this strategy ensures that the rates of various enforcement activities being compared across officers have the same denominators [15, 29]. We term this the "*common circumstances*" assumption. (For a detailed discussion of potential threats to this assumption, see SI Appendix C.3). To further enhance the credibility of the common-circumstances assumption, we additionally subset to individuals for whom personnel data could be merged. Only those of rank "police officer" are retained, eliminating sergeants and other higher-ranked officers that may not patrol as regularly. Non-standard shifts are dropped (retaining only first through third watches), as are absences, shifts with assigned special duties (e.g. station security or training),

[8] In analyzing officer behavior, we take patrol assignments—the level at which commanding officers typically exercise control—as the basic unit of analysis. In SI Section C.4, we examine the possibility that differences in clock-in/out times may be one mechanism behind observed differences in behavior between officer racial groups. While some statistically significant differences can be observed (roughly 0.1% disparities in patrol time), these gaps are two orders of magnitude smaller than the differences observed in stops, arrests, and uses of force that make up our main results.

and unusual double- or triple-duty days in which a single officer serves multiple shifts. Events occurring while the officer was off duty are also eliminated. SI Appendix A.1 describes these datasets and our preprocessing procedures in detail.

For ease of interpretation, 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 SI Appendix C.2 for a discussion of estimation and SI Figures A8-A10 for these additional results). All statistical inferences are based on officer-level block bootstrap confidence intervals that are robust to unobserved officer-specific peculiarities. Our behavioral analysis organizes patrol assignments into over 650,000 groups defined by MDSBs. Of these, roughly 300,000 allow for comparisons between officers of differing racial/ethnic groups (and thus contribute to our analyses); 230,000 contain both female and male officers; and 50,000 contain Hispanic officers of differing language abilities. For details, see SI Appendix C.5.

Importantly, our assignment records allow tracking of officer-shifts *whether or not they engaged in any enforcement activity*. Here, our analysis differs from most prior studies of officer race and gender using police administrative data [e.g. 3, 9, 20, 34, 44], which only rely on instances in which officers recorded an activity (e.g. a stop or arrest; see discussion in [22]). (See [19], [45], and [46] for other strategies to combat selection bias.) Without these patrol assignment records, inaction is invisible to the analyst, and officers who do not record any enforcement activity simply vanish from the data. In addition, as we show below, officers of different races and ethnicities are assigned to work in systematically different conditions. Patrol assignment data allows us to make apples-to-apples comparisons by adjusting for this fact.[9]

We caution that our approach is uninformative about causal mechanisms. While we observe that white and nonwhite officers behave differently under common circumstances, we cannot discern the psychological pathways or other channels through which these differences arise. But results nonetheless demonstrate the efficacy of proposed personnel reforms: they show *what behavior civilians can expect from police when officers of a given demographic profile are deployed, on average*, holding environmental factors fixed. While we believe investigation of such mechanisms are an important direction for future work, given the urgent need for evidence on police reform efficacy, careful estimation of overall effects is our paramount concern.

[9] Our data indicate that two officers were often involved in stopping or arresting a civilian. Throughout, such instances enter as separate rows in our officer-shift data, since in such cases each officer contributed to the stop/arrest. However, to gauge robustness, SI Appendix C.7 reanalyzes stops and arrests after restricting the data to a single officer per event, which yields substantively identical results.

4 Results

We now turn to whether deploying officers of different demographic profiles affects the volume and nature of police-civilian interactions. Table 3 displays the average differences in the number of stops, arrests and uses of force associated with Black and Hispanic officers (relative to white officers) and female officers (relative to male) working in the same MDSBs. We also note that roughly 77% of Hispanic residents in the Chicago metropolitan area speak Spanish at home as of 2015 [23]. However, our personnel data show less than half of Hispanic CPD officers can speak Spanish. Because of the potential for this language barrier to affect police-civilian interactions within this ethnic group, we therefore also compare Hispanic officers who can and cannot speak Spanish.

Turning first to Black officers, we see that when faced with the same working conditions, this group makes 1.52 fewer stops per 10 shifts, makes 1.94 fewer arrests per 100 shifts, and uses force 1.02 fewer times per 1,000 shifts, on average, than white counterparts—that is, compared to white officers assigned to patrol the same beat, in the same month, on the same day of the week, and at the same shift time (all $p_{\text{adj}} < 0.001$ after Benjamini-Hochberg multiple-testing correction for comparisons reported in Table 3). These gaps are large, representing reductions of 29%, 21%, and 32% relative to typical stop, arrest, and use-of-force volume for white officers (see SI Table A1 for average behavior by officer race).

Importantly, Table 3 also shows these disparities are not uniform across situations. Rather, they are driven by a reduced focus on engaging Black civilians (1.26 fewer stops per 10 shifts and 1.46 fewer arrests per 100 shifts, 39% and 25% of typical white officer behavior) and a broad class of enforcement activities that can be thought of as relatively discretionary, including stops of civilians for "suspicious behavior" (-0.57 per 10 shifts, 31% less) or arrests for drug-related crimes (-0.31 per 100 shifts, 27% less). However, when it comes to enforcing violent crime, Black officer behavior looks similar to that of white officers, with Black officers making only 0.22 fewer arrests per 100 shifts for violent crimes than whites (only 11% less). In other words, when it comes to policing the most serious offenses, Black officers are roughly comparable to white officers. Black officers also deploy force against Black civilians 0.85 fewer times per 1,000 shifts (38% less) than their white counterparts. (All adjusted p-values < 0.001.) In fact, reduced use of force against this civilian group accounts for 83% of the overall force disparity between white and Black officers. (SI Section C.6 and SI Figures A8–A10 show that results are virtually identical using a wide range of alternative estimators.) This pattern of results is remarkably in line with the hopes of proponents of racial diversification, who seek to reduce abusive policing and mass incarceration in Black communities.

We also find substantial differences in the behavior of female officers relative to male officers.

Relative to male officers working on the same places and times, female officers make 0.61 fewer arrests total per 100 shifts (-7% relative to typical male behavior) and 0.54 fewer arrests of Black civilians per 100 shifts (-9%, both $p_{\text{adj.}} < 0.001$). In fact, 89% of this disparity in arrest rate is due to reduced arrests of Black civilians. We also find that female officers use force 0.87 less times overall (-28%) and 0.71 fewer times per 1,000 shifts against Black civilians (-31%, both $p_{\text{adj.}} < 0.001$), with the latter accounting for 82% of overall force reduction.

Table 2: **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 88 hypotheses: * is < 0.05, ** is < 0.01, and *** is < 0.001. Alternative estimators, including adjusting for officer experience, are reported in SI Figures A8–A10 and produce substantively identical results.

Stops per 10 shifts

	Civilian race			Reason					
	Black	Hispanic	White	Other	Loitering	Suspicious	Drug	Traffic	Total
Black officers (vs white)	-1.26^{***}	-0.13^{***}	-0.13^{***}	-0.25^{***}	-0.03^{***}	-0.57^{***}	-0.17^{***}	-0.50^{***}	-1.52^{***}
Hispanic officers (vs white)	-0.29^{***}	0.05	-0.04^{*}	-0.02	0.00	-0.17^{***}	-0.05^{***}	-0.05	-0.28^{***}
Female officers (vs male)	-0.03	-0.01	0.02	0.16^{***}	-0.01	-0.16^{***}	-0.07^{***}	0.07	-0.01
Spanish-speaking Hispanic officers (vs non-Spanish-speaking)	0.03	0.01	-0.01	0.07	-0.02	-0.06	-0.03	0.08	0.04

Arrests per 100 shifts

	Civilian race:			Reason:					
	Black	Hispanic	White	Other	Drug	Traffic	Property	Violent	Total
Black officers (vs white)	-1.46^{***}	-0.29^{***}	-0.18^{***}	-0.93^{***}	-0.31^{***}	-0.18^{***}	-0.29^{***}	-0.23^{***}	-1.94^{***}
Hispanic officers (vs white)	-0.30^{*}	-0.08	-0.08^{*}	-0.32^{***}	0.03	-0.08^{**}	-0.08^{*}	0.02	-0.44^{**}
Female officers (vs male)	-0.54^{***}	-0.09	0.02	-0.28^{***}	-0.17^{***}	-0.02	-0.02	-0.12^{**}	-0.61^{***}
Spanish-speaking Hispanic officers (vs non-Spanish-speaking)	-0.48	-0.20	-0.04	-0.17	-0.14	-0.13	-0.19^{*}	-0.12	-0.74^{*}

Uses of force per 1,000 shifts

	Civilian race			
	Black	Hispanic	White	Total
Black officers (vs white)	-0.85^{***}	-0.10	-0.05	-1.02^{***}
Hispanic officers (vs white)	-0.33^{***}	0.03	-0.06	-0.37^{*}
Female officers (vs male)	-0.71^{***}	-0.13^{**}	-0.00	-0.87^{***}
Spanish-speaking Hispanic officers (vs non-Spanish-speaking)	-0.02	0.20	0.21	0.38

Table 3: **Within each racial/ethnic group, officers of differing genders behave differently when facing common circumstances.** Each row reports a comparison between female and male officers of a particular racial/ethnic group. Each column contains average per-shift differences in a police behavior among officers assigned to the same beat and shift time, during the same day of week and month. Asterisks indicate Benjamini-Hochberg p-values accounting for officer-level clustering and adjusted for multiple testing of 66 hypotheses: * is < 0.05, ** is < 0.01, and *** is < 0.001.

Stops per 10 shifts (gender differences within racial/ethnic group)

	Civilian race			Reason					
	Black	Hispanic	White	Other	Loitering	Suspicious	Drug	Traffic	Total
Black female officers (vs male)	0.19	0.02	0.02	0.22***	-0.00	-0.06	-0.04***	0.12	0.23
Hispanic female officers (vs male)	-0.09	-0.01	0.04	0.11	-0.02	-0.16	-0.03	0.05	-0.03
White female officers (vs male)	-0.03	-0.02	-0.00	0.11	-0.00	-0.16*	-0.05	0.07	-0.04

Arrests per 100 shifts (gender differences within racial/ethnic group)

	Civilian race:			Reason:					
	Black	Hispanic	White	Other	Drug	Traffic	Property	Violent	Total
Black female officers (vs male)	-0.33	-0.01	0.00	-0.18	-0.05	0.03	-0.01	-0.13	-0.34
Hispanic female officers (vs male)	-0.62*	-0.05	-0.01	-0.15	-0.20	-0.15	-0.10	-0.09	-0.69
White female officers (vs male)	-0.40	-0.09	0.07	-0.24	-0.14	-0.01	0.04	-0.09	-0.44

Uses of force per 1,000 shifts (gender differences within racial/ethnic group)

	Civilian race			
	Black	Hispanic	White	Total
Black female officers (vs male)	-0.52*	-0.01	-0.02	-0.54*
Hispanic female officers (vs male)	-1.13***	-0.37	-0.02	-1.52***
White female officers (vs male)	-0.61*	-0.19	-0.00	-0.87***

Results differ in important ways for Hispanic officers. Like their Black colleagues, Hispanic officers facing the same working conditions conduct fewer stops, arrests and uses of force than white officers. However, these differences are far more modest. Strikingly, gaps are primarily driven by less engagement with Black civilians, while Hispanic officers exhibit nearly the same volume of enforcement activity against co-ethnic civilians as white officers, on average. Our estimates indicate that Hispanic officers can be expected to make 0.28 fewer stops per 10 shifts (6% reduction relative to typical white officer behavior, $p_{\text{adj.}} < 0.001$) and 0.37 fewer uses of force per 1,000 shifts ($p_{\text{adj.}} = 0.017$, 12% reduction) per shift. As the table shows, part of this heterogeneity is driven by Spanish-speaking Hispanic officers, who make 0.74 fewer arrests per 100 shifts on average than Hispanic officers who do not speak the native language of many co-ethnic communities (6% less, $p_{\text{adj.}} = 0.038$). When it comes to stops and uses of force, differences in the behavior Spanish and non-Spanish-speaking Hispanic officers appear negligible. However, we caution that these results are in part due to the relatively small sizes of these language groups, which dramatically reduce overlap between Spanish- and non-Spanish-speaking Hispanic officers in MDSBs. An improved understanding of language and other cultural factors remains an important direction for future work.

5 Discussion and Conclusion

Fatal encounters between white police officers and unarmed racial minorities continue to prompt widespread calls for law enforcement reforms. Personnel reforms are prominent among these proposals, but evidence on their efficacy has been unclear due to severe data limitations. Using an unusually rich individual-level data set on police activity in Chicago, we provide the most credible micro-level assessment to date of these reforms' impact on police-civilian interactions.

Our results reveal the efficacy of officer diversity, while also underscoring several previously unknown nuances and avenues worthy of further exploration. Faced with the same working conditions, 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 a reduced focus on the enforcement of discretionary stops and arrests for petty crimes, including drug offenses, which have long been thought to fuel mass incarceration [2]. In contrast to these drastic differences in the policing of petty crime, Black officers' enforcement of violent crime is only slightly lower than that of white officers.

But this pattern, which closely comports with the hopes of advocates of racial diversification, is complicated by several additional results. For one, while Hispanic officers display lower levels of enforcement activity than whites overall, their behavior toward Hispanic civilians is broadly comparable to that of white officers. However, we find evidence that after accounting for language ability, Spanish-speaking Hispanic officers make fewer arrests than their non-Spanish-speaking counterparts. 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 these gendered reductions are driven by a reduced focus by female officers on arresting and using force against Black civilians.

Our analysis has important limitations. First, we analyze data from a single city, a choice which allows for an unusually rich analysis but comes at the potential expense of external validity. We hope that our analysis provides a useful template for other scholars to test whether these patterns hold in other places and times as additional similar data becomes available. A second limitation is that we cannot identify specific mechanisms connecting officer race and/or gender to behavior. The effects we demonstrate directly inform the efficacy of racial diversification policies, but the mechanisms behind these effects deserve further study.

In addition, while our analysis tests for tangible effects of diversification, we emphasize that a large literature on descriptive representation [4, 41, 13] suggests diversifying police agencies may yield intangible benefits, namely, increased trust in government among historically underrepresented groups [43]. If racial diversity affects levels of public trust, the cost-benefit calculus when considering this class of reforms would be further complicated. As agencies diversify, institutional culture may also begin to shift and exert its own effects on officer behavior—an important long-run implication that is beyond the scope of our study.

Taken together, our results show that the effects of racial diversification are neither simple nor monolithic. Like the civilians they serve, officers are multidimensional. Crafting effective personnel reforms requires thinking beyond the coarse demographic categories typically used in diversity initiatives, and consideration of how multiple officer attributes relate police to the civilians they serve.

References

[1] Adams, Megan Marie. 2012. The Patrolmen's Revolt: Chicago Police and the Labor and Urban Crises of the Late Twentieth Century PhD thesis UC Berkeley.
URL: *https://escholarship.org/uc/item/9gp045vr*

[2] Alexander, Michelle. 2010. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness.* New York: The New Press.

[3] Antonovics, Kate and Brian G. Knight. 2009. "A new look at racial profiling: Evidence from the Boston Police Department." *The Review of Economics and Statistics* 91(1).

[4] Arnold, R. Douglas. 1990. *The Logic of Congressional Action.* Yale University Press.

[5] Ba, Bocar A. 2019. "Going the Extra Mile: the Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago." *Working paper* .

[6] Byrne, John, Rick Pearson and Jeremy Gorner. 2018. "Mayor Rahm Emanuel rejects President Donald Trump's call for Chicago police to use stop-and-frisk tactics." *Working paper* . `https://www.chicagotribune.com/nation-world/ct-trump-chicago-shootings-20181008-story.html`.

[7] DOJ/EEOC. 2016. "Advancing Diversity in Law Enforcement.". `https://www.justice.gov/crt/case-document/file/900761/download`.

[8] Donohue III, John J. and Steven D. Levitt. 2001. "The impact of race on policing and arrests." *The Journal of Law and Economics* 44(2):367–394.

[9] Fagan, Jeffrey, Anthony A. Braga Rod K. Brunson and April Pattavina. 2016. "Stops and stares: Street stops, surveillance, and race in the new policing." *Fordham Urb. LJ* 43:539–614.

[10] Fantz, Ashley. 2020. "Want to reform the police? Hire more women." *CNN* . `https://www.cnn.com/2020/06/23/us/protests-police-reform-women-policing-invs/index.html`.

[11] Forman Jr., James. 2017. *Locking up our own: Crime and punishment in Black America.* Farrar, Straus and Giroux.

[12] Garner, Maryah, Anna Harvey and Hunter Johnson. 2019. "Estimating Effects of Affirmative Action in Policing: A Replication and Extension." *International Review of Law and Economics* p. 105881.

[13] Gay, Claudine. 2002. "Spirals of trust? The effect of descriptive representation on the relationship between citizens and their government." *American Journal of Political Science* pp. 717–732.

[14] Gelman, Andrew, Jeffrey Fagan and Alex Kiss. 2007. "An Analysis of the New York City Police Department's "Stop-and-Frisk'" Policy in the Context of Claims of Racial Bias." *Journal of the American Statistical Association* 102(429):813–823.

[15] Goff, Phillip Atiba and Kimberly Barsamian Kahn. 2012. "Racial bias in policing: Why we know less than we should." *Social Issues and Policy Review* 6(1):177–210.

[16] Harvey, Anna and Taylor Mattia. 2019. "Reducing Racial Disparities in Crime Victimization." *Working Paper* . `https://s18798.pcdn.co/annaharvey/wp-content/uploads/sites/6417/2019/12/Victimization_Harvey_Mattia.pdf`.

[17] Heckman, James J. 1977. "Sample selection bias as a specification error (with an application to the estimation of labor supply functions)." *NBER Working Paper* (No. 172).

[18] Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–161.
URL: *http://www.jstor.org/stable/1912352*

[19] Hoekstra, M. and C. Sloan. 2020. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." *National Bureau of Economic Research, Working Paper (26774)* . `https://econrails.uzh.ch/api/agenda/event_attachments/269`.

[20] Hoffman, Peter B. and Edward R. Hickey. 2005. "Use of force by female police officers." *Journal of Criminal Justice* 33(2):145–151.

[21] Knox, Dean and Jonathan Mummolo. 2020. "Making Inferences About Racial Disparities in Police Violence." *Proceedings of the National Academy of Sciences* 117:1261–1262.

[22] Knox, Dean, Will Lowe and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* . `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=33363387`.

[23] Krogstad, Jens Manuel and Mark Hugo Lopez. 2017. "Use of Spanish declines among Latinos in major U.S. metros." *Pew Research Center* . `https://www.pewresearch.org/fact-tank/2017/10/31/use-of-spanish-declines-among-latinos-in-major-u-s-metros/`.

[24] Legewie, Joscha and Jeffrey Fagan. 2016. "Group threat, police officer diversity and the deadly use of police force." *Columbia Public Law Research Paper* . `http://raceandpolicing.issuelab.org/resources/25209/25209.pdf`.

[25] Lerman, Amy and Vesla Weaver. 2014. *Arresting Citizenship: The Democratic Consequences of American Crime Control.* University of Chicago Press.

[26] Lonsway, Kim, Michelle Wood Megan Fickling Alexandria De Leon M. Moore P. Harrington E. Smeal and Katherine. Spillar. 2002. "Men, women, and police excessive force: A tale of two genders.". `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=33363387`.

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

[28] Miller, Amalia R and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *The Review of Economic Studies* 86(5):2220–2247.

[29] Neil, Roland and Christopher Winship. 2019. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." *Annual Review of Criminology* 2(1):73–98.
URL: *https://doi.org/10.1146/annurev-criminol-011518-024731*

[30] *Power, Politics, & Pride: Afro-American Patrolmen's League*. 2018. Library Catalog: interactive.wttw.com.
URL: *https://interactive.wttw.com/dusable-to-obama/afro-american-patrolmens-league*

[31] Rabe-Hemp, Cara E. 2008. "Female officers and the ethic of care: Does officer gender impact police behaviors?" *Journal of Criminal Justice* 36(5):426–434.

[32] Raganella, Anthony J. and Michael D. White. 2004. "Race, gender, and motivation for becoming a police officer: Implications for building a representative police department." *Journal of Criminal Justice* 32(6):501–513.

[33] Reichel, Philip L. 1988. "Southern slave patrols as a transitional police type." *American Journal of Police* 7:51–77. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=33363387`.

[34] Ridgeway, Greg and John M. MacDonald. 2009. "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops." *Journal of the American Statistical Association* 104:661–668.

[35] Rim, Nayoung, Roman G. Rivera and Bocar A. Ba. 2019. "In-group Bias and the Police: Evidence from Award Nominations." *Working paper* .

[36] Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and non-randomized studies." *Journal of Educational Psychology* 66:688–701.

[37] Rubin, Donald B. 1990. "Formal mode of statistical inference for causal effects." *Journal of statistical planning and inference* 25(3):279–292.

[38] Sen, Maya and Omar Wasow. 2016. "Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics." *Annual Review of Political Science* 19:499–522.

[39] Sklansky, David Alan. 2005. "Not your father's police department: Making sense of the new demographics of law enforcement." *J. Crim. L. & Criminology* 96:1209–1244.

[40] Smith, Brad. 2003. "The Impact of Police Officer Diversity on Police-Caused Homicides." *Policy Studies* 31(2):147–162.

[41] Tate, Katherine. 2001. "The political representation of blacks in Congress: Does race matter?" *Legislative Studies Quarterly* pp. 623–638.

[42] Turner, K. B., David Giacopassi and Margaret Vandiver. 2006. "Ignoring the past: Coverage of slavery and slave patrols in criminal justice texts." *Journal of Criminal Justice Education* 17(1):181–195.

[43] Tyler, T.R. and J. Fagan. 2008. "Legitimacy and cooperation: Why do people help the police fight crime in their communities?" *Ohio State Journal of Criminal Law* 6(1):231–275.

[44] Weisburst, EK. 2019. "Police use of force as an extension of arrests: Examining disparities across civilian and officer race." *American Economic Asssociation: Papers and Proceedings* 109.

[45] Weisburst, Emily. 2020. "'Whose help is on the way?' The importance of individual police officers in law enforcement outcomes." *Working Paper* . `https://drive.google.com/file/d/1z04Hrq1Muv9z0oTuFRWddtp8Pbg7m4aa/view`.

[46] West, Jeremy. 2018. "Racial Bias in Police Investigations.". Working Paper `https://people.ucsc.edu/~jwest1/articles/West_RacialBiasPolice.pdf`.

[47] White, Michael D., Jonathon A. Cooper Jessica Saunders and Anthony J. Raganella. 2010. "Motivations for becoming a police officer: Re-assessing officer attitudes and job satisfaction after six years on the street." *Journal of Criminal Justice* 38(4):520–530.

Supplemental Information

Table of Contents

A Detailed Description of Data

A.1 CPD Data

The administrative data from the CPD used in this study span multiple data sets 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 and officers' home address in 2004, 2005 and in early 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 data set 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 33,000 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).

A.2 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 data sets, 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.

A.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.

A.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 "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.

A.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. station security or training), and unusual 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."

B Descriptive Analysis of Assignment Patterns

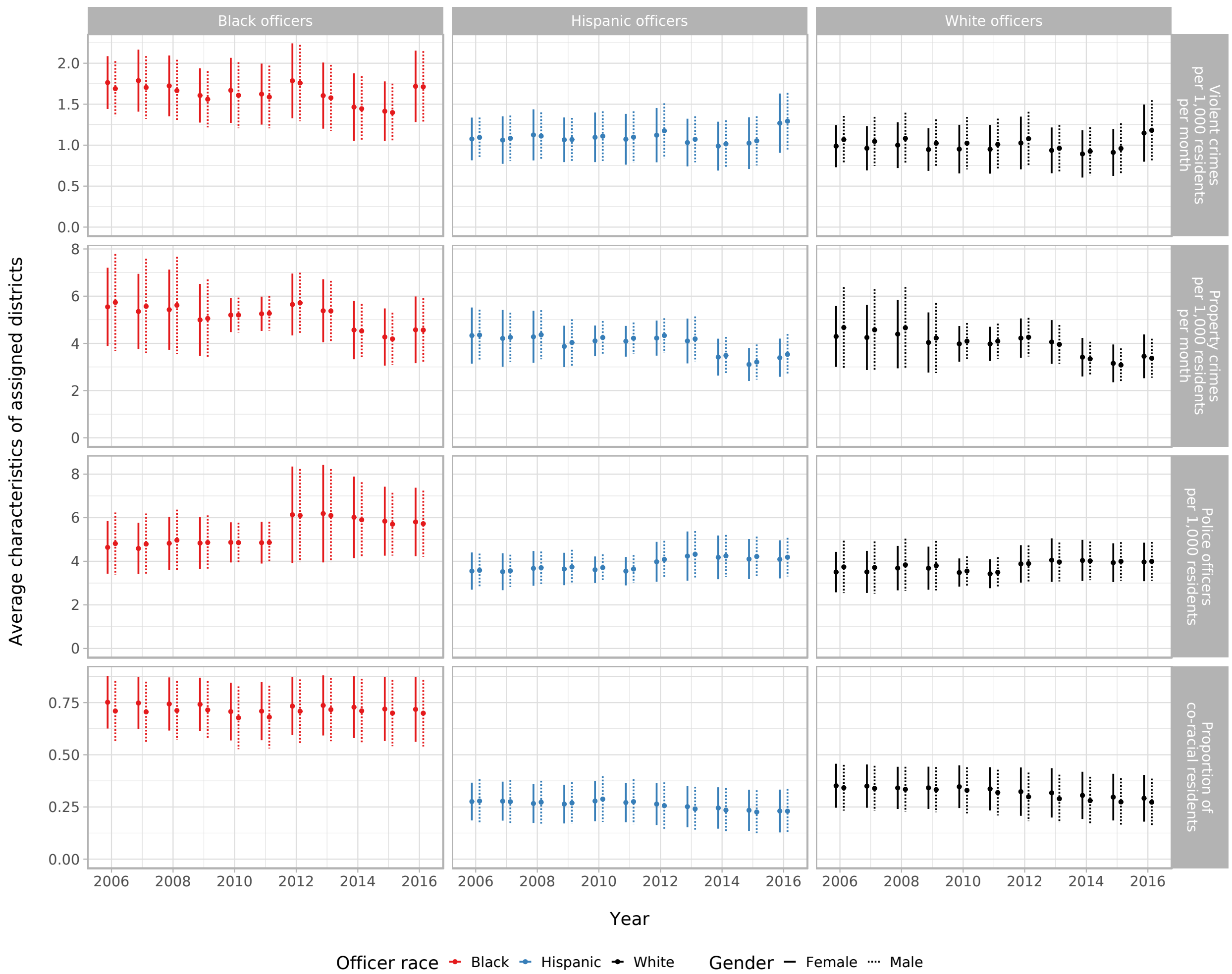
B.1 District Characteristics

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[1]) 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 A1 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

[1] See Chicago Police Department 2008 Annual Report, page 37, https://home.chicagopolice.org/wp-content/uploads/2014/12/2008-Annual-Report.pdf

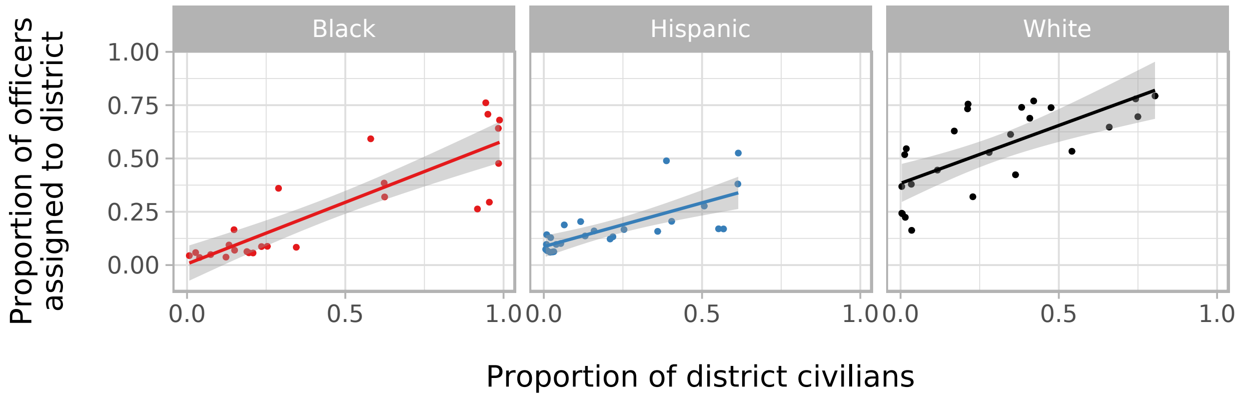
Figure A1: Average characteristics of assigned geographic districts for various officer groups, by year, based on 1,089,707 monthly assignments to geographic units from 2006–2016. Confidence intervals cluster on district-month.



proportion of co-racial residents. (These results are analogous to those presented in Figure 2, but break out each year separately rather than pooling.) Results are highly comparable, indicating that reported patterns are not an artifact of pooling across years.

We now turn to two district-level analyses. Figure A2 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 3) is 98% Black, but 24% of officers assigned there are white.

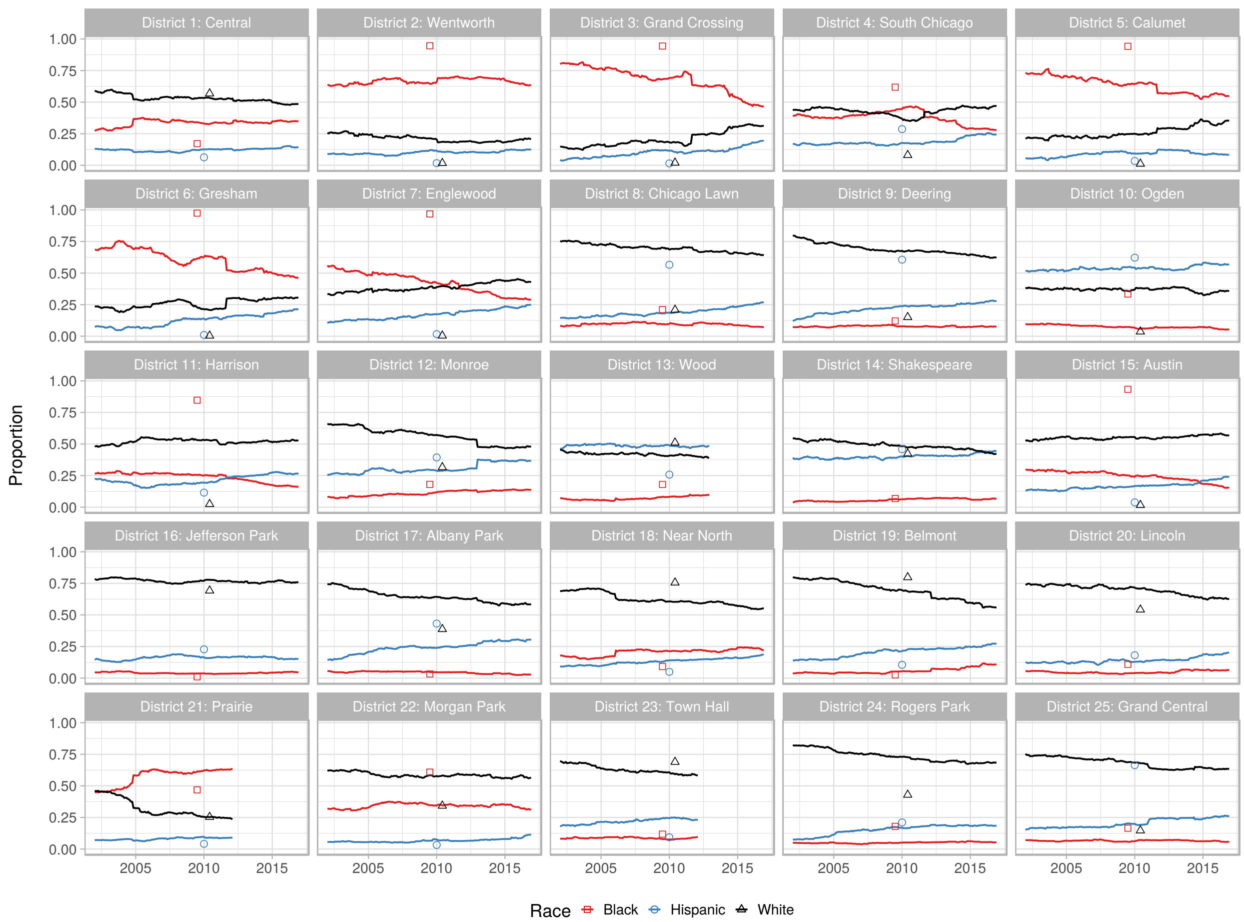
Figure A2: **Racial and ethnic composition of officers' assigned districts.** In each panel, each point represents a police district. The horizontal axis indicates the proportion of civilians of a given racial/ethnic group residing in 2010 Census data, and the vertical axis depicts the share of officers assigned to that district in January 2010 from the same racial/ethnic group.



The disparity is even starker in Austin (CPD District 15), where a 96% Black resident population is policed by a unit that is 55% white. (See SI Section A.3 for details on the computation of resident demographics.)

Figure A3 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 A2. 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.

Figure A3: **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.



B.2 Officer Race 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 A4–A5 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 A6–A7 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 3) 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.

Figure A4: **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).

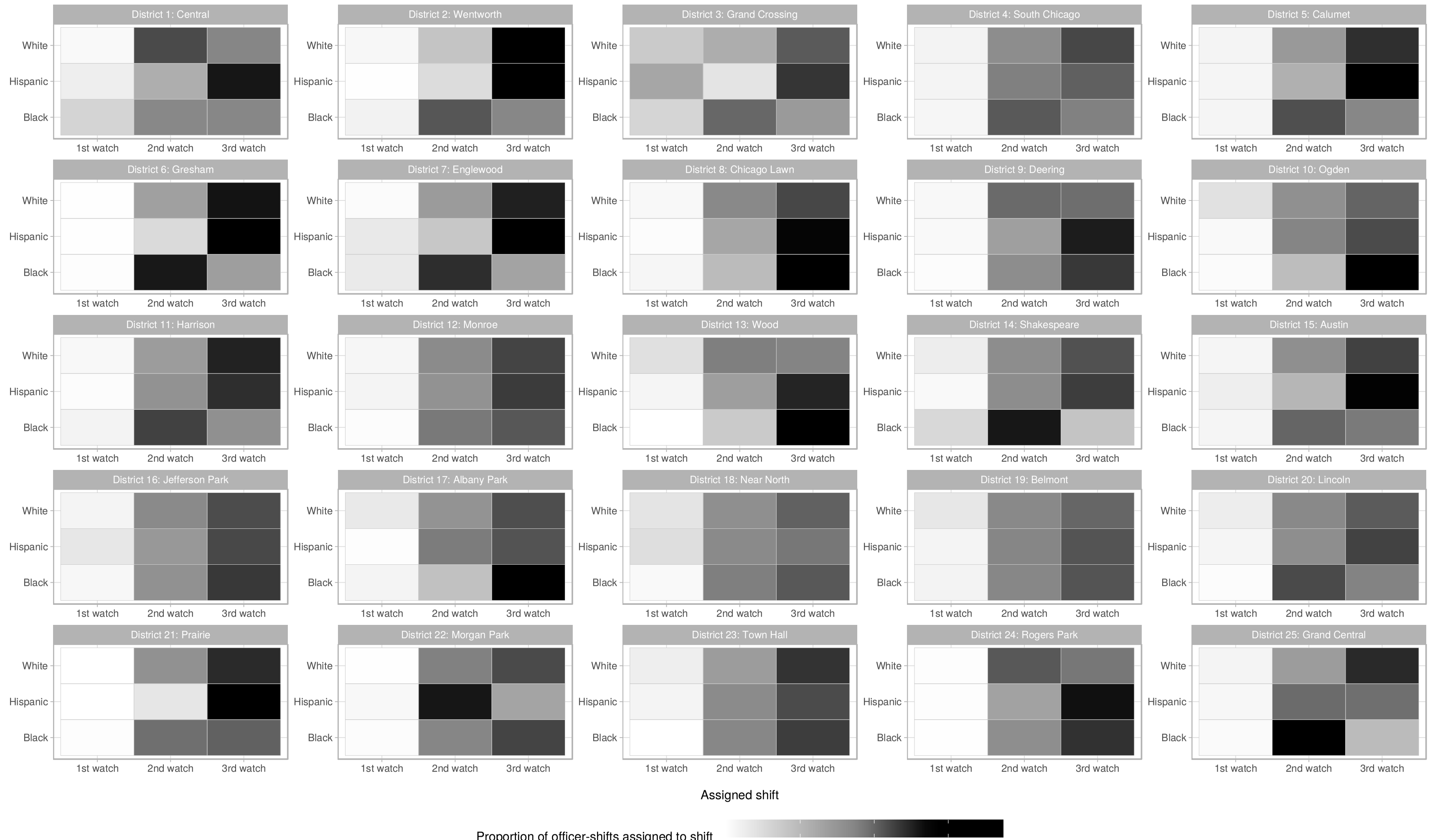


Figure A5: **Assigned shift time by officer gender.** Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for female or male officers to first watch (midnight to 8 a.m.), second watch (8 a.m. to 4 p.m.), and third watch (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).

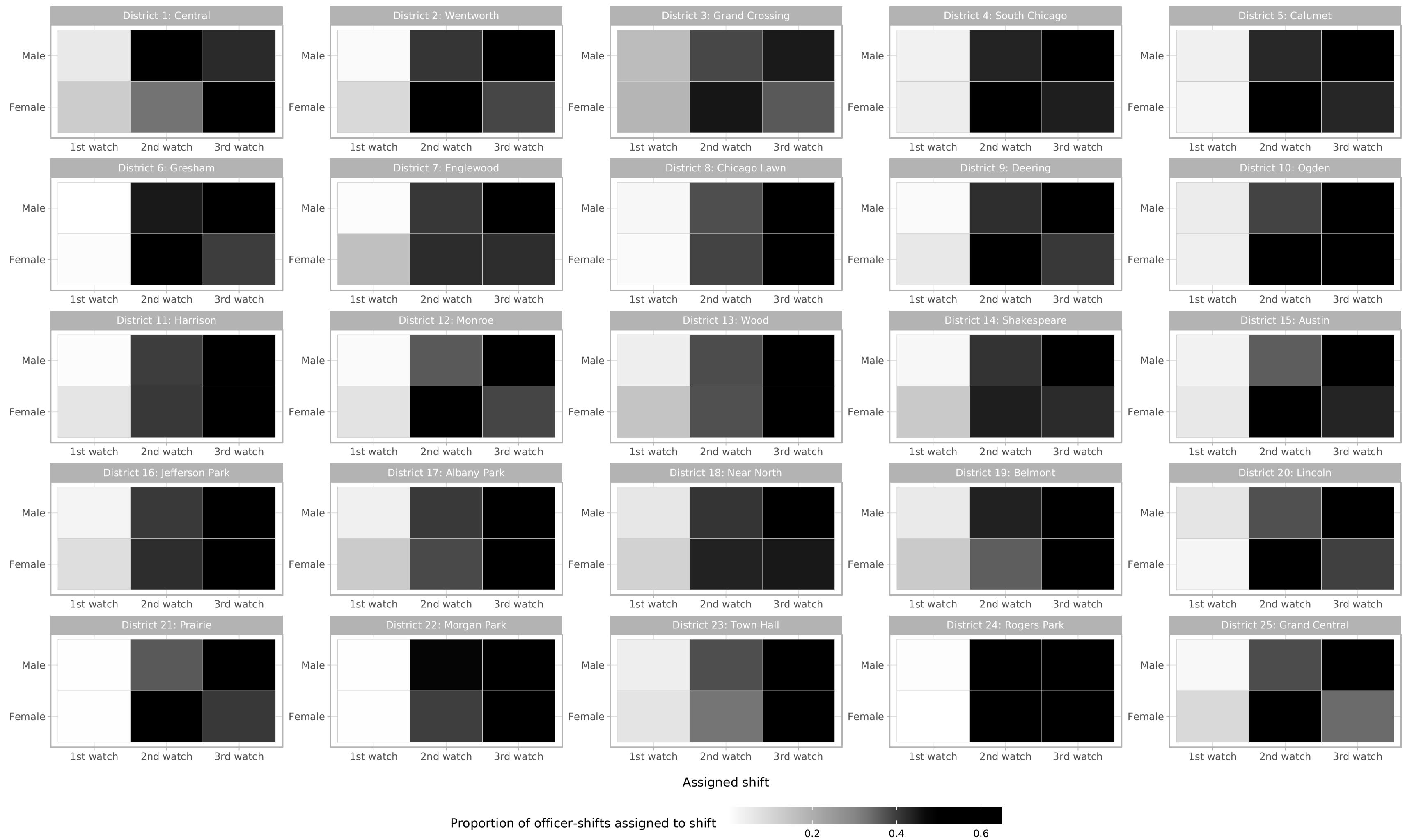


Figure A6: **Assigned beat by officer race.** Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for Black, Hispanic, or white officers to beats, or geographic patrol areas. Darker cells indicate a higher proportion of 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).

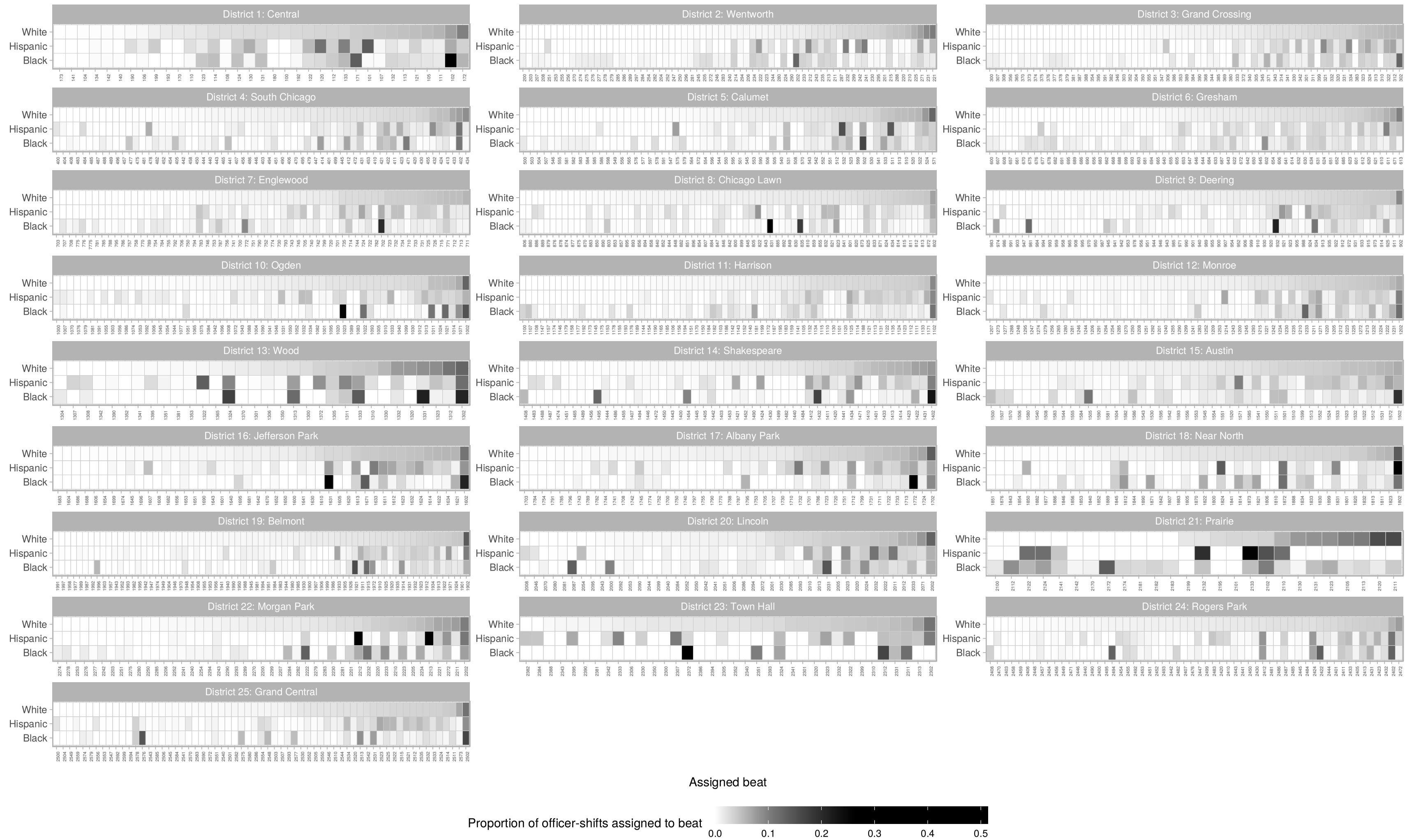
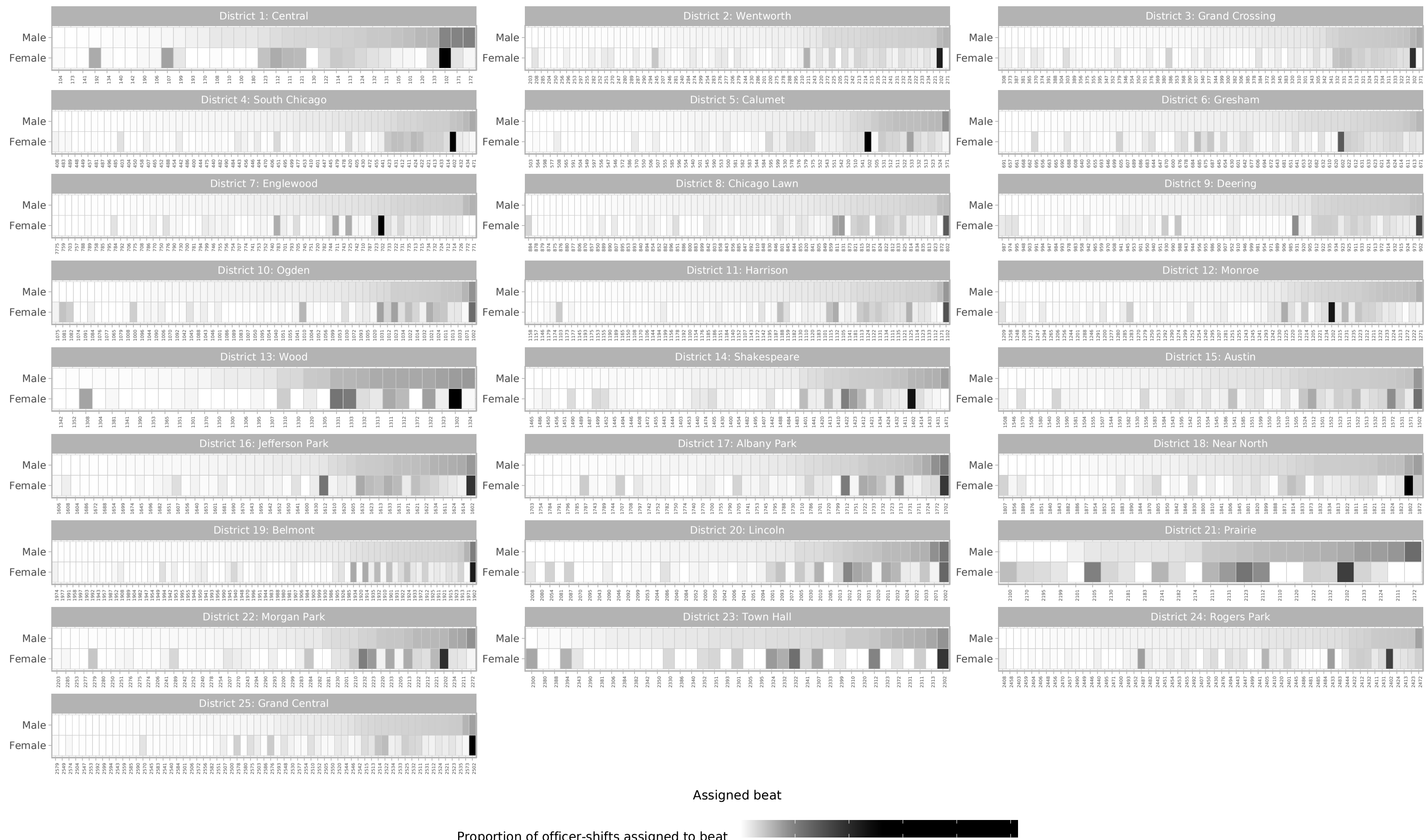


Figure A7: **Assigned beat by officer gender.** Each panel depicts officers within a geographic police district. Within a district, each row shows the proportion of shift assignments for female or male officers to beats, or geographic patrol areas. Darker cells indicate a higher proportion of assignments 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).



C Officer Behavior

C.1 Unadjusted Average Behavior by Various Officer Groups

Table A1: **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 10 shifts:				
Civilian race: Black	3.03	2.61	3.09	3.25
Civilian race: Hispanic	0.92	0.21	1.43	1.10
Civilian race: White	0.57	0.21	0.61	0.75
Reason: Drug	0.48	0.18	0.63	0.58
Reason: Loitering	0.06	0.03	0.09	0.07
Reason: Other	1.31	1.16	1.35	1.39
Reason: Suspicious	1.54	0.89	1.73	1.84
Reason: Traffic	1.18	0.81	1.39	1.30
Total:	4.58	3.06	5.19	5.17
Arrests per 100 shifts:				
Civilian race: Black	5.64	4.95	5.93	5.91
Civilian race: Hispanic	1.76	0.40	2.52	2.19
Civilian race: White	0.89	0.32	0.98	1.17
Reason: Traffic	0.70	0.35	0.78	0.87
Reason: Drug	0.94	0.36	1.24	1.13
Reason: Other	3.10	2.04	3.46	3.54
Reason: Property	1.52	1.15	1.66	1.68
Reason: Violent	2.11	1.82	2.38	2.16
Total:	8.37	5.71	9.52	9.36
Force per 1,000 shifts:				
Civilian race: Black	1.99	1.45	2.10	2.26
Civilian race: Hispanic	0.41	0.09	0.58	0.50
Civilian race: White	0.26	0.07	0.28	0.36
Total:	2.71	1.63	3.02	3.19

Table A2: **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 10 shifts:			
Civilian race: Black	3.03	2.39	3.25
Civilian race: Hispanic	0.92	0.65	1.02
Civilian race: White	0.57	0.49	0.59
Reason: Drug	0.48	0.27	0.55
Reason: Loitering	0.06	0.04	0.07
Reason: Other	1.31	1.23	1.34
Reason: Suspicious	1.54	1.06	1.71
Reason: Traffic	1.18	0.99	1.25
Total:	4.58	3.59	4.92
Arrests per 100 shifts:			
Civilian race: Black	5.64	4.03	6.18
Civilian race: Hispanic	1.76	1.11	1.98
Civilian race: White	0.89	0.69	0.95
Reason: Traffic	0.70	0.45	0.79
Reason: Drug	0.94	0.47	1.09
Reason: Other	3.10	2.09	3.44
Reason: Property	1.52	1.21	1.63
Reason: Violent	2.11	1.68	2.26
Total:	8.37	5.90	9.20
Force per 1,000 shifts:			
Civilian race: Black	1.99	1.10	2.29
Civilian race: Hispanic	0.41	0.20	0.48
Civilian race: White	0.26	0.20	0.28
Total:	2.71	1.52	3.11

Table A3: **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 (Spanish Hisp. off.)	Mean (Non-Spanish Hisp. off.)
Stops per 10 shifts:		
Civilian race: Black	2.15	4.03
Civilian race: Hispanic	1.50	1.37
Civilian race: White	0.65	0.59
Reason: Drug	0.53	0.72
Reason: Loitering	0.06	0.11
Reason: Other	1.25	1.47
Reason: Suspicious	1.47	2.01
Reason: Traffic	1.05	1.74
Total:	4.36	6.05
Arrests per 100 shifts:		
Civilian race: Black	3.99	7.84
Civilian race: Hispanic	2.26	2.81
Civilian race: White	0.98	0.99
Reason: Traffic	0.46	1.10
Reason: Drug	0.72	1.76
Reason: Other	2.54	4.38
Reason: Property	1.55	1.80
Reason: Violent	2.06	2.69
Total:	7.33	11.73
Force per 1,000 shifts:		
Civilian race: Black	1.35	2.80
Civilian race: Hispanic	0.52	0.66
Civilian race: White	0.30	0.28
Total:	2.21	3.79

C.2 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 (geographic patrol area) 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. 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 [SUTVA, 37], 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.[2]

The slot-level policy effect is the difference in potential outcomes [36] $Y_{i,j}(r) - Y_{i,j}(r')$, 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,

[2] We explore the validity of this second assumption to the extent possible in SI Appendix C.7, 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.

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 3, a key identifying assumption is that

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

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 nonwhite 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 $Y_{i,j} = \alpha_i + \sum_r \beta_r \mathbf{1}(R_{ij} = r)$, where α_i represents a fixed effect for MDSB i. Unbiasedness of this estimator requires the additional assumption that MDSB-specific policy effects are homogeneous, or that $\delta_i = \delta_{\mathcal{R}}$ 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 C.6. As we show in SI Figures A8–A10, these alternative results are virtually identical to our primary results.

C.3 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. 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 causal estimates relating to officer *deployment*. (For related discussions of conceptualizing race as a causal variable, see [21] and [38]).

C.4 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, 85.5% are 9 hours in duration, with 8.5- and 8-hour shifts making up an additional 7.8% and 5.1%, respectively (percentages are based on rounding shift duration to the nearest 6 minutes). In fixed-effect regression analyses that compare officers within unique MDSB combinations, we estimate that shifts of Black officers are 0.0094 hours shorter (roughly 0.1% shorter) than their white counterparts assigned to the same MDSB, and Hispanic officer shift durations are statistically indistinguishable from 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.

C.5 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 653,527 unique combinations of month, day of week, shift number, and beat (MDSBs). Of these, 572,067 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 watchs 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 294,963 MDSBs that contain overlap between multiple assigned officer racial/ethnic groups (e.g., one Black officer and one white officer); 229,143 MDSBs contain overlap between both female and male officers; and 49,721 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.

C.6 Alternative Estimators

Our primary analysis of officer behavior uses OLS regression with MDSB fixed effects, of the form $Y_{i,j} = \alpha_i + \sum_r \beta_r \mathbf{1}(R_{ij} = r)$, where α_i represents a fixed effect for MDSB i. As we discuss in SI Section C.2, 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 A8–A10 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 C.2, 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.

Figure A8: **Stops.** 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.

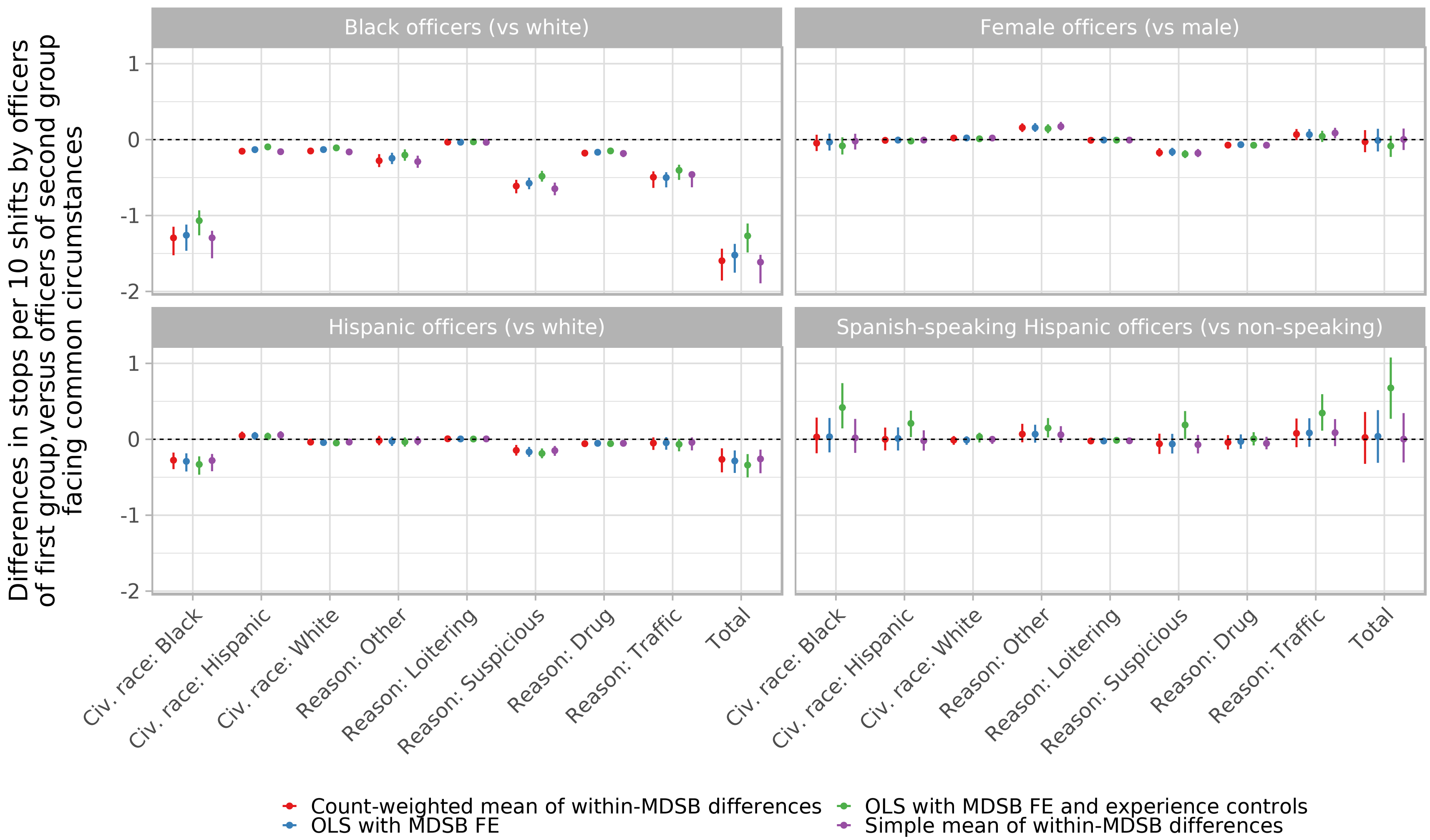
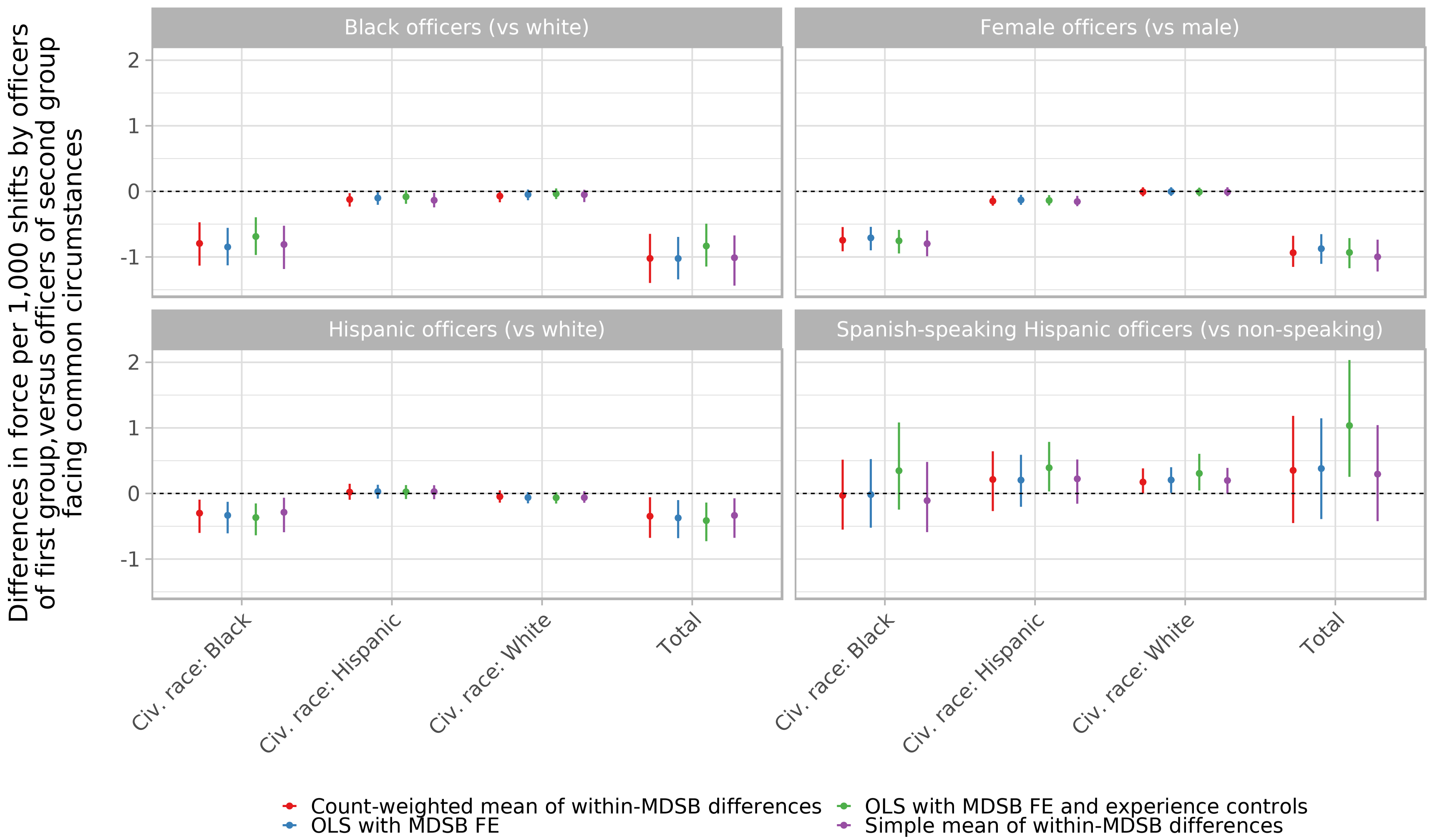


Figure A9: **Arrests.** Points (error bars) depict estimated differences (95% block-bootstrap confidence intervals) in the number of arrests conducted by Black and Hispanic officers per shift, relative to white officers patrolling in the same month, day group (weekday/weekend), shift time (first/second/third watch), and beat. Results presented using four different estimators.

Figure A10: **Uses of Force.** 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.



C.7 Robustness Checks: Multiple Stopping Officers

Data on stops and arrests of civilians indicate that the vast majority of both activities are jointly reported by two officers. 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 and arrests in which we use only data on first reporting officers and a randomly drawn officer from each event, respectively. Results are substantively unchanged. Note that a small number of results lose significance—such as the reduction in female-officer drug stops (versus male officers) and Black-officer arrests of white civilians (versus white officers)—and a handful of other comparisons appear to become marginally significant. However, general patterns (and all findings reported in the main text) remain substantively identical, with smaller coefficients reflecting the fact that roughly half of all stop and arrest events have been discarded.

Table A4: **Robustness of demographic differences to analyzing one stopping/arresting officer per encounter.**. Following Table 3, 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 88 hypotheses: * is < 0.05, ** is < 0.01, and *** is < 0.001. Results are substantively identical to Table 3.

Stops per 10 shifts

	Civilian race			Reason					
	Black	Hispanic	White	Other	Loitering	Suspicious	Drug	Traffic	Total
Black officers (vs white)	-0.88***	-0.11***	-0.10***	-0.21***	-0.02***	-0.42***	-0.12***	-0.32***	-1.09***
Female officers (vs male)	0.12	0.04	0.03*	0.19***	-0.00	-0.10***	-0.01	0.12***	0.20*
Hispanic officers (vs white)	-0.34***	-0.01	-0.05***	-0.05	0.01	-0.20***	-0.05***	-0.10*	-0.40***
Spanish-speaking Hispanic officers (vs non-Spanish-speaking)	-0.01	-0.08	0.01	0.06	-0.04	-0.04	-0.13***	0.06	-0.08

Arrests per 100 shifts

	Civilian race:			Reason:					
	Black	Hispanic	White	Other	Drug	Traffic	Property	Violent	Total
Black officers (vs white)	-0.72***	-0.16***	-0.08	-0.45***	-0.17***	-0.09***	-0.14***	-0.11***	-0.97***
Female officers (vs male)	-0.28***	-0.04	0.01	-0.16***	-0.08***	-0.01	-0.00	-0.07	-0.32***
Hispanic officers (vs white)	-0.14*	-0.06	-0.03	-0.17***	-0.01	-0.03	-0.04	0.02	-0.24***
Spanish-speaking Hispanic officers (vs non-Spanish-speaking)	-0.32***	-0.16	0.01	-0.14	-0.11	-0.08	-0.05	-0.11	-0.48*