American Economic Review 2022, 112(3): 827–860
https://doi.org/10.1257/aer.20201292

Does Race Matter for Police Use of Force? Evidence from 911 Calls[†]

By Mark Hoekstra and CarlyWill Sloan*

This paper examines race and police use of force using data on 1.6 million 911 calls in two cities, neither of which allows for discretion in officer dispatch. Results indicate White officers increase force much more than minority officers when dispatched to more minority neighborhoods. Estimates indicate Black (Hispanic) civilians are 55 (75) percent more likely to experience any force, and five times as likely to experience a police shooting, compared to if White officers scaled up force similarly to minority officers. Additionally, 14 percent of White officers use excess force in Black neighborhoods relative to our statistical benchmark. (*JEL* H76, J15, K42, R23)

There are widespread concerns regarding police officer treatment of minorities. These concerns are rooted in a long history of police mistreatment of Black Americans, and are reflected by the fact more Black civilians report having "no confidence" in local police (24 percent) than have "a lot of confidence" (14 percent). This distrust of police is perhaps strongest with respect to police use of force, as only 33 percent of Black civilians believe officers use the right amount of force for the situation, and only 35 percent believe police treat racial and ethnic groups equally (Pew Research Center 2016). Concerns over the role of race in police use of force have been voiced most forcefully in the protests over police shootings of unarmed Black civilians and by the Black Lives Matter movement. Most recently, this movement has grown exponentially in the aftermath of the George Floyd murder, with more than 1,700 demonstrations across all 50 states (Haseman et al. 2020). Beyond the direct harms caused by inequitable police use of force, the distrust this engenders likely reduces policing efficiency and increases the social costs of crime.

However, documenting whether race matters for police use of force is difficult. This is in part because researchers often do not observe interactions in which force

* Sloan: Claremont Graduate University (email: carlywill.sloan@cgu.edu); Hoekstra: Texas A&M University, NBER, and IZA (email: markhoekstra@tamu.edu). Stefano DellaVigna was the coeditor for this article. We would like to thank the police departments in two anonymous cities for providing the data. We would also like to thank multiple officials in those cities for helpfully answering our numerous questions. We also thank Isabelle Sin, Scott Carrell, Jan Feld, Felipe Goncalves, Benjamin Hansen, Kevin Lang, Samuel Myers, Jr., James West, seminar participants at Boston University, University of Florida, Georgia Tech, University of Mississippi, University of Maryland, Mississippi State University, Notre Dame, Haverford College, National University of Singapore, and conference participants at the 2019 Stata Texas Empirical Microeconomics Conference, the Victoria University of Wellington Applied Econometrics Workshop, the 2019 SEA Annual Conference, the 2020 ASSA Annual Conference, and the 2020 NBER Summer Institute for helpful comments and suggestions. This paper was completed under Texas A&M IRB 2018-0712D. Any errors are our own.

[†] Go to https://doi.org/10.1257/aer.20201292 to visit the article page for additional materials and author disclosure statements.

was not used. As a result, researchers must make assumptions regarding the appropriate "benchmark," such as violent crime rates or arrests. In addition, it is hard for researchers to observe whether the underlying risk of situations involving White and minority civilians, or White and minority officers, merit similar force. It is at best unclear whether controlling for observed contextual factors is sufficient to overcome bias due to selection. This is because officers almost always observe factors not recorded in the data, much of what is in the data is recorded by police after force was used, and the recorded characteristics of the encounter could themselves be impacted by race.

In this paper, we use a fundamentally different approach we believe most closely mirrors the ideal thought experiment. Specifically, we observe settings in which White and minority officers are as-good-as-randomly sent to otherwise similar situations, and where the same Black and White officers are observed responding to situations in White and minority neighborhoods. We do this using administrative data on 1.6 million calls to 911, which enables us to observe police use of force for a defined set of interactions, independent of whether the interaction involved use of force. Importantly, the data come from two cities in which the dispatch protocols allow for no discretion on the part of the officer or the operator with respect to which officer is dispatched. Rather, in the first city, the dispatcher observes on the computer screen whether the officer on duty for that beat is available and if so, dispatches that officer. If the dispatcher's monitor indicates the beat officer is unavailable—which happens if, for example, the beat officer is currently engaged in a traffic stop—then she dispatches the available officer observed to be closest to the call's location. The protocol for the second city requires that the operator dispatch the available officer who is closest geographically to the location. Both protocols imply that conditional on police beat and time fixed effects, the variation in the race of the officer dispatched is as good as random. Our interviews with dispatchers indicate they follow the protocol. We also show empirically that officer race in both cities is uncorrelated with call characteristics and with predicted use of force based on those covariates.

We use this exogenous variation in police officer race to answer two questions. The first is whether White officers use force at higher rates than minority officers when responding to otherwise similar calls. The second is whether officers are more likely to use force on different-race civilians. We answer this second question by asking whether White officers increase their use of force more than minority officers as they are dispatched to more-minority neighborhoods. We do so in a difference-in-difference style framework similar to the Price and Wolfers (2010) study on racial bias by NBA referees, except that we focus primarily on specifications that include officer fixed effects to control for any sorting of officers across neighborhoods. The advantage of this approach is we avoid imposing selection-on-observables assumptions in comparing interactions across civilian race. Rather, we assume that in the absence of an different-race effect, White and minority officers should increase use of force by a similar amount as they are dispatched from White to minority neighborhoods. The limitation is that as in all use of force studies, we do not observe the proper amount of force. As a result, our measure of whether civilian race matters is necessarily a relative one: how would Black civilians have fared if the White officers had scaled up their use of force the same as Black officers?

Our data include the universe of 911 calls made in two cities linked to police officer race and use of force. As a condition of acquiring the data, we cannot disclose the names of the cities. The first city has a population and police force composed primarily of White and Black officers. It has a population of more than 240,000 and a homicide rate that ranks in the top 20 among the nation's largest 100 cities. In this city, we have administrative records on over 1.2 million calls to 911 over a three-to-five-year period starting after 2010.[1] These calls resulted in 1,300 police uses of force, 94 of which involved the discharge of an officer's gun. The data include the time and date of the call, the priority score assigned to the call by the 911 operator, a short description of the call, the first officer(s) dispatched to the scene by the operator, and whether or not force was used (and by which officer) at the scene. Importantly, we observe the first officer(s) dispatched to the scene, even if other officers also arrive after the dispatched officer arrives. In addition, we observe the address from which the call originated, which we geocode into a census block group to assign civilian race. In the second city, we have data on over 400,000 calls to 911. This city is composed primarily of White and Hispanic civilians and police officers. This city has a population of more than 150,000, and the call records include similar information. There are 389 incidents of police use of force linked to these calls, to which 663 officers were dispatched. We do not observe the type of force used in this city, or which officer used force.

Results indicate that White officers use force 60 percent more often than Black officers on average, and use force with a gun more than twice as often. In both cases, this difference is driven primarily by White officers using much higher rates of force in Black neighborhoods compared to Black officers. Perhaps most strikingly, we show that while White and Black officers use force with a gun at approximately the same rate in White and racially mixed neighborhoods, White officers use force with a gun five times as often in neighborhoods that are at least 80 percent Black. This difference is driven entirely by within-officer differences; estimates from officer fixed effects specifications indicate dispatching a White officer to a call in a Black neighborhood increases the use of gun force by five times. Estimates for overall force indicate that White officers scale up use of force 55 percent more than Black officers as they go to Black neighborhoods. Similarly, results from the second city indicate that even though White and Hispanic officers use force at the same overall rate, use of force is disproportionately concentrated in different-race neighborhoods. Estimates indicate that dispatching an officer of a different race increases use of force by 75 percent. Across both cities, we show that effects are robust to controlling for the beat with which the officer is most familiar, or that beat interacted with officer race. We also show that our results are not explained by White officers responding differently to observed call characteristics, or by less experienced or more male officers responding differently to calls in minority neighborhoods.

We interpret our findings as providing strong evidence that race matters in police use of force. Importantly, we remain agnostic about the mechanism underlying the effects. These mechanisms include racial animus by the officer, as well as statistical discrimination based on false beliefs—e.g., officers overestimating risk in

[1] We do not report the exact years to protect the anonymity of the city.

different-race neighborhoods. This interpretation is consistent with a large experimental literature in social psychology in which subjects reveal bias against Black civilians when deciding whether to shoot (e.g., see Mekawi and Bresin 2015 for a meta-analysis). However, estimates also pick up the effect of differential skill. For example, minority officers may be equally adept at de-escalating situations in White neighborhoods, but better at doing so in minority ones. Finally, while estimates necessarily capture civilian response endogenous to officer behavior—which we consider to be part of the officer behavior channel—they also pick up the net impact of civilian response to officer race per se. If civilians respond more aggressively to different-race versus same-race officers who act the same way, such behavior could explain at least some of our difference-in-difference estimates. On the other hand, if civilians respond less aggressively to different-race officers—perhaps out of fear—then adjusting for this would increase our estimates. We view these responses as both possible and unobservable, and certainly as empirically indistinguishable from a response caused by officer behavior.

We also perform an analysis of individual officers in the first city by estimating (shrunken) individual officer effects. We use these officer effects to flag individuals who are further in the tails of the use of force distribution than one would expect, even given the number of officers in the sample. This analysis, which replicates Ridgeway and MacDonald's (2009) implementation of Efron's (2004) method, flags 14 percent of White officers, versus 4 percent of Black officers, as being far in the right tail of the use of force distribution in Black neighborhoods. Additionally, we show that if one were to try to reduce disparities by eliminating officers in the higher use-of-force group, one would have to eliminate 32 percent of White officers in the right tail to eliminate the overall difference between White and Black officers, and 31 percent of White officers to reduce the different-race effect to zero.

The main contribution of this paper is to overcome the benchmarking problem by using 911 calls, and to overcome the selection problem by studying contexts where White and minority officers are exogenously assigned to calls in White and minority neighborhoods. In doing so, this paper joins a larger literature examining the impact of race in the criminal justice system. This includes studies that examine differences between the propensity of White and Black officers in firing their guns at incidents with multiple officers present (Ridgeway 2016) and in making traffic stops in similar neighborhoods (Ba et al. 2021). It is related to work on racial bias in police vehicle searches (e.g., Anwar and Fang 2006, Persico and Todd 2006, Antonovics and Knight 2009). To address the issue of officer selection into interactions with civilians, this literature tests for racial bias by modeling police behavior and implementing tests based on vehicle search "hit rates" relative to a benchmark encounter rate. A related literature addresses the difficulty of assessing the benchmark encounter rate by exploiting changes in ambient light to test for racial profiling in traffic stops (e.g., Grogger and Ridgeway 2006, Horrace and Rohlin 2016). Most recently, Goncalves and Mello (2021) cleverly use the behavior of nonlenient officers to measure racial bias in writing traffic tickets. This paper complements these literatures by taking a substantively different approach to solving problems created by endogenous police-civilian interactions. In doing so, this paper is more closely related to work by Weisburst (2017), who uses 911 calls to estimate the value added of individual police officers, and West (2018), who tests for racial bias in traffic citations using conditionally random variation in the race of officers called to traffic accidents. The advantage of this paper relative to West (2018) is we examine the impact of race in arguably a more important context with more important outcomes—911 calls and police use of force. The advantage of West (2018) is he has objective information on whether certain citations were merited, which is not possible for use of force. In addition, this paper also complements research on racial bias in the criminal justice system more generally, including racial bias by prosecutors (Rehavi and Starr 2014, Sloan 2019, Tuttle 2019), juries (Anwar, Bayer, and Hjalmarsson 2012; Flanagan 2018) and judges (Arnold, Dobbie, and Yang 2018; Bielen, Marneffe, and Mocan 2018).[2]

A second contribution of our paper is to demonstrate how the 911 call dispatch system can be used to assess the performance of individual officers, as well as for officers grouped by characteristics such as race. The major advantage of doing so in this context is to overcome the problems encountered elsewhere due to nonrandom officer selection into situations. This approach also has the advantage of being straightforward to implement with existing data in other cities. Many cities likely have dispatch protocols in place already that would enable a simple yet compelling analysis, and with modest changes to the protocol every city could.

In assessing the effect of officer and civilian race on use of force, this paper is most similar to work by Fryer (2019) and Johnson et al. (2019). Fryer (2019) uses an impressive range of datasets including detailed data on all interactions from stop-and-frisk in New York City. In addition, he uses data on officer-involved shootings in Houston, which he benchmarks using arrests. Using these data, he implements a selection-on-observables design to control for contextual factors. He concludes that Blacks and Hispanics are more likely to experience nonlethal force all else equal, but not more likely to experience an officer-involved shooting. Weisburst (2019) extends this work using data on use of force and arrests from Dallas, and similarly reports that conditional on arrest there is no racial difference in use of force. Johnson et al. (2019) use data on fatal officer-involved shootings across counties and conclude there is little evidence of bias, though others have criticized the underlying assumptions in the analysis (Knox and Mummolo 2019; Knox, Lowe, and Mummolo 2019). The main advantage of this paper relative to this prior work is we are able to estimate the effect of race in a context where Black and White officers are as-good-as-randomly dispatched to similar situations, and where each officer is as-good-as-randomly dispatched to calls in more and less Black neighborhoods. In this way, we avoid making potentially problematic assumptions about the appropriate benchmark. In addition, by assessing the impact of civilian race by comparing how White and Black officers scale up use of force as they are dispatched from White to Black neighborhoods, we avoid concerns about whether we control for enough contextual factors that may differ across civilian race. Similarly, we avoid potentially overcontrolling for factors described by the officer after force was used (e.g., in an arrest report), which would lead to understating effects. The limitations

[2] More generally, it also complements a broader literature on own-group bias in education, labor, housing, and product markets (e.g., Feld, Slamanca, and Hamermesh 2016; Ayres and Siegelman 1995; Dahl and Moretti 2008; Goldin and Rouse 2000; Lavy 2008; Neumark, Bank, and van Nort 1996; Moss-Racusin et al. 2012; Price and Wolfers 2010; Parsons et al. 2011.)

of our approach are that our difference-in-difference estimate of whether race matters is necessarily a relative one, and we address this question in only two cities.

Our results have important implications for policing in the United States. Perhaps most importantly, they provide rigorous evidence supporting the common civilian perception that race is an important determinant of police use of force. The results of this paper suggest that at least in the contexts studied here, this belief seems warranted, especially with respect to the use of lethal force. In addition, this study demonstrates that race matters even in a time and context during which police departments generally, and individual officers in particular, know they are under close scrutiny by the media and the public.

I. Background and Data

A. *Background and Dispatch Procedures*

As noted above, the protocol for dispatching officers to the scene of 911 calls is critical for our research design. For this reason, we contacted police departments in more than a dozen cities inquiring about their system for dispatching officers to calls, as well as the availability of data. In particular, we needed to be able to observe and link the race of the police officer to 911 calls and use of force. We were able to obtain data in two cities that met both criteria.[3] As part of the agreement to obtain the necessary data, which includes officer identifiers, we were required not to disclose the names of the cities. However, we can state that the first city we study has large Black and White populations, a total population of over 240,000 and has a homicide rate that ranks in the top 20 among the nation's 100 largest cities. We note this contrasts with the cities studied by Fryer (2019) and Weisburst (2019), none of which have a homicide rate that ranks in the top 20 among the nation's largest cities.

In this city, a civilian's 911 call is given to the first dispatcher available. The dispatcher then records important aspects of the call. Specifically, the computer system used by dispatchers records the time, exact location, and police beat of the call. The dispatcher will then ask the caller about details surrounding the call, categorize the seriousness and urgency of the call and rate it from highest priority (one) to low priority (higher values). The dispatcher also records a short description of the call. For example, a dispatcher may record a call as a "domestic disturbance"and then assign it a priority of two. Calls are then dispatched based on the priority of the call. This means more urgent calls, like assaults or crimes in progress, will be dispatched first, while less serious calls, like stolen cars, will be given lower priority and dispatched later.

After recording the above aspects of the call, the dispatcher assigns a primary unit to the call. In the first city, one officer is assigned as the on-duty officer for each beat during each shift, and is expected to patrol that beat if not dispatched to a call for service. Protocol dictates that the dispatcher first dispatches the beat officer for that call's beat, if available. An officer will not be available if he is responding to

[3] We also obtained data from a third city that we were told used a protocol in which officers and operators had no discretion. However, upon receiving the data we discovered that did not appear to be the case, and thus do not use that city in this analysis.

another call for service or self-initiated event.[4] For example, if an officer makes a traffic stop, he will use his in-car computer to communicate that he is not available, in which case he will not appear on the police dispatcher's screen for that period of time. If the beat officer is not available, the geographically closest officer will be dispatched. This is determined by the dispatcher, who observes each available officer's location on screen. We spoke with multiple dispatchers in this city, who confirmed this process and the lack of officer or dispatcher discretion in which officer to dispatch. The majority of calls (98 percent) in the first city are assigned only one primary unit. In 1.2 percent of calls, two primary units are dispatched, while in the remaining calls there are three to five primary units. We note that in both cities, after the primary unit(s) is dispatched, other officers may observe the call on their police car computer and respond. We do not observe these officers in our data, and assign only the primary dispatched officer(s) to each call. In this way, we perform an intent-to-treat analysis.

This dispatch procedure leads directly to our research design. Specifically, we rely on controlling for either beat fixed effects or for beat-by-year-by-week-by-shift fixed effects to isolate the as-good-as-random variation in officer race.[5] In this way our design reflects the following thought experiment: Much of the time the beat officer, who may be Black or White, will be dispatched to a call. But some of the time, that officer will be unavailable. As a result, the next-closest officer will be dispatched, who may be of a different race. In Section III, we show empirical evidence consistent with the identifying assumption that conditional on beat or beat-by-time fixed effects, call characteristics and risk are uncorrelated with officer race. In addition, we are able to address the potential threat to identification that (say) White officers are less familiar with more dangerous Black neighborhoods where (say) Black officers are assigned as beat officers. We note this does not seem to be an issue in practice; the average neighborhood race for calls responded to by Black and White officers are 60.3 and 57.5 percent Black, respectively, and the home beats for Black and White officers are 57.6 and 55.1 percent Black, respectively. Regardless, we show results when we control directly for a fixed effect for the beat where an officer responds to the most calls, and (in difference-in-difference specifications) for an interaction between home beat and officer race.

Once a primary unit is dispatched to a call they may encounter a situation that leads to use of force. If an officer uses any type of force, police department administrative procedure dictates they must immediately file a report describing the details of the incident and the use of force type. This report number will be recorded in the officer in-car computer and linked to the call for service where the use of force occurred. Even if the use of force report is made later—for example, after the officer has been dispatched to another call—we are still able to link this use of force to the call for service using the police report number. A use of force report must be recorded even in events where nondeadly force (punches/kicks, etc.) is used. In the event an officer discharges his gun, he must allow the ranking officer on the

[4] These data do not include officer-initiated incidents where officers observe an incident, call it in, and have the dispatcher assign them to that incident.

[5] Shifts are staggered in this city. We implement this fixed effect by defining shifts as 8 AM–4 PM, 4 PM–midnight, and midnight–8 AM.

scene to inspect the weapon and issue replacement ammunition. If an officer has shot someone, a detective or internal affairs investigator will take possession of the weapon. All use of force reports are reviewed quarterly by a community use of force committee, which makes recommendations about the use of force policy to the chief of police.

We also study use of force in a second city, where the population is more than 95 percent Hispanic or White, and less than 10 percent Black. This city has a population of more than 150,000, which ranks in the largest 300 cities in the country. The protocol for dispatching officers to calls is similar to the city described above in that calls are dispatched according to a protocol that does not allow for discretion on the part of the operator or officer. However, in this city, the operator first dispatches the geographically closest available officer to the call. In addition, the biggest difference in the dispatch process in this city is that multiple officers are dispatched more frequently. Specifically, 44 percent of calls have one unit assigned, 31 percent have two units assigned, and 25 percent have three or more units assigned. As with the first city, after any use of force, department procedure dictates that the officer record the incident report electronically, and that their supervisor review the report.

B. *Data*

The police administrative dataset for the first city includes all calls for service from a three-to-five-year period after 2010 where at least one officer was dispatched. For each call in our dataset, we observe the primary unit, beat, priority, time between call and dispatch, latitude, longitude, time of the call, time of dispatch, and date of the call. There are over 50 beats in this city, and three to four times as many census block groups.[6] We also observe the race, gender, and years of experience of each police officer. Because Hispanics make up less than 5 percent of police officers in our city, we exclude those officers from the sample. Additionally, we observe if the call resulted in use of force and the type of use of force. Type of force is recorded by the police officer, including whether the force involved the firing of a gun by the officer. We classify use of force in two different ways; results are nearly identical. We report results where force is defined at the call level in our main analysis, and report results where force is defined at the officer level in online Appendix Table A6.[7]

In the first city, we define civilian race as the proportion of the population that is Black in the census block group from which the call originated. We do so using the 2010 census for each of the several hundred different census block groups from which calls in our city originated (United States Census Bureau 2010).[8] In the second city, we assign civilian race as the proportion of the population that is minority

[6] In order to protect the anonymity of the cities, we are not revealing the exact time period or number of beats.

[7] In the first approach, we classify all officers in the assigned primary unit(s) as having used force if any officer assigned to the call used force. For example, if two officers are dispatched to the same event, but only one uses force, in this approach we assign both officers as having used force. We use this event-level assignment procedure in order to account for the joint decision-making process of responding officers. For example, one officer could escalate a situation, causing another dispatched officer to use force. The second way we classify use of force is to assign it at the officer level, in which an officer is only classified as having used force if that officer used force.

[8] We classify civilians as Black if they are only Black, and White if they are only White. This results in the classification of over 90 percent of this city's citizens as only Black or only White. Hispanics make up less than 5 percent of the population in this city.

(e.g., Black or Hispanic), noting that less than 5 percent of the population in that city is Black. Importantly, we note that any measurement error generated by using neighborhood race as a proxy for civilian race should be uncorrelated with the race of the police officer, given how the dispatch process works.[9]

As with many cities in the United States, there is significant sorting by race across neighborhoods. For example, in the first city about 35 percent of our census block groups are more than 75 percent Black and about 25 percent of our census block groups are more than 75 percent White. This is also evident in the distribution of 911 calls. This distribution is shown in Figure 1, where proportion Black for the census block group of the originating call is shown on the x-axis and varies from zero to one. It is clear that while there are calls originating from all types of neighborhoods, we have a significant number of calls originating from nearly-all-White or (especially) nearly-all-Black neighborhoods. Panels B and C of Figure 1 show that Black and White officers are dispatched to both types of neighborhoods, as well as neighborhoods of mixed race. In addition, in Appendix Figure A2, we show there is significant within-officer variation in neighborhood race across calls. Specifically, we first estimate the standard deviation in proportion Black of the neighborhood of the call attended for each officer. We then graph the distribution of standard deviations by officer race. Online Appendix Figure A2 shows that the mass of both distributions is between 0.2 and 0.35. This suggests most officers are dispatched across neighborhoods that vary widely in racial composition.

In the first city, we observe a total of 1,233,139 officer-call observations representing a total of 1,225,521 calls. There were 1,341 incidents of force committed by 600 different officers, and 94 incidents of use of force with a gun committed by 68 different officers. We note that the rarity of use of force, and especially use of force with a gun, necessarily presents challenges for statistical power. However, we show later that estimated differences are still largely significant at conventional levels using a variety of methods of statistical inference.

Summary statistics are shown in Table 1, where officer-call observations are weighted by the inverse of the number of officers dispatched to a call. Table 1 shows that force is used just over 1 in 1,000 calls, while use of force with a gun is used in just fewer than 1 in 10,000 calls (0.0076 percent). Overall, 7 percent of use of force involved a gun, 38 percent involved a taser, and the remaining 55 percent was grouped into a category that included use of hands, feet, mace, baton, etc. to subdue the civilian. Thirty-eight percent of responding officers were Black, and 16 percent were female. Average officer experience was ten years. The average racial composition across neighborhoods was 58.6 percent Black. It takes 6.5 minutes for a primary unit to be dispatched to a call.

In columns 2 and 3 we show summary statistics separately for Black and White responding officers, respectively. We note that this comparison does not reflect our research design since it does not account for potential officer selection by race into different police beats. However, given these data only include use of force resulting from 911 calls, and since officers of both races respond to calls in all neighborhoods

[9] In addition, we believe the most likely type of measurement error would lead us to understate effects. For example, if it were true that some civilians with whom police interacted even in all White neighborhoods were Black, then our estimates would need to be scaled up to capture the actual change in civilian race across neighborhoods.

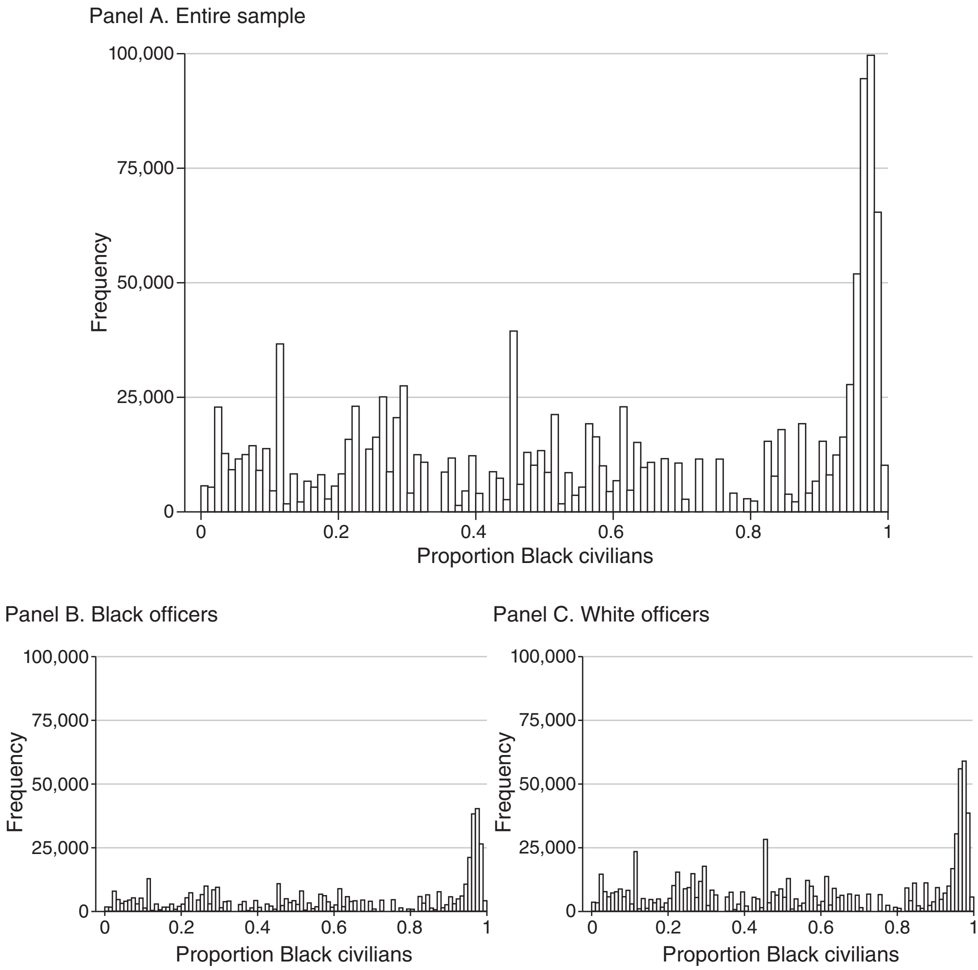

Figure 1. Distribution of 911 calls across Census Block Groups

Notes: These figures report the distribution of proportion Black civilians for 911 calls. Each histogram uses 0.01 size bins. Panels B and C report the histograms for calls where Black or White officers are dispatched, respectively.

at roughly the same rates (as shown in Figure 1) and have no discretion in how they are dispatched, we find it instructive nonetheless. Columns 2 and 3 show that Black officers are somewhat more likely to be female (18.9 versus 14.2 percent), have slightly more experience (10.5 versus 9.9 years), and respond to calls in slightly more Black neighborhoods (60.3 versus 57.5 percent). Black and White officers respond to calls of similar priority and are dispatched similarly quickly (6.51 versus 6.46 minutes) and to locations with similar x and y coordinates. However, use of force is quite different across officer race. While Black officers use force 7.8 out of every 10,000 calls, White officers use force 12.8 times per 10,000 calls. White officers use force with a gun approximately 1 out of every 10,000 calls, which is more than twice as often as Black officers.

Summary statistics for the second city are shown in online Appendix Table A8, where we observe 938,562 officer-call observations representing 414,633 calls. Eighty-five percent of census block groups are minority, and 81 percent are Hispanic.

Table 1—Summary Statistics

	Entire sample (1)	Black officers (2)	White officers (3)
Outcomes			
Use of force	0.00106	0.000764	0.00124
Gun use of force	0.0000710	0.0000397	0.0000903
Call characteristics			
Proportion Black civilians	0.586	0.603	0.575
	(0.333)	(0.333)	(0.333)
Per capita income	23,281.7	22,793.1	23,582.0
	(14,844.8)	(14,550.4)	(15,015.1)
Proportion unemployed	0.139	0.143	0.137
	(0.111)	(0.112)	(0.110)
Proportion less than HS degree	0.185	0.187	0.184
	(0.118)	(0.117)	(0.118)
Priority of call	2.839	2.839	2.838
	(0.757)	(0.754)	(0.758)
Time between call and dispatch	6.490	6.519	6.473
	(63.43)	(23.10)	(78.54)
X coordinate	87.30	87.31	87.30
	(0.808)	(0.788)	(0.820)
Y coordinate	111.6	111.7	111.6
	(1.685)	(1.672)	(1.690)
Hour	13.33	13.00	13.53
	(6.496)	(6.651)	(6.391)
Call from home beat	0.180	0.183	0.178
Black officer	0.381	1	0
Female officer	0.160	0.189	0.142
Years of experience	10.10	10.52	9.841
	(7.972)	(7.965)	(7.966)
Observations	1,233,139	469,170	763,969
Number of calls	1,225,521	466,560	758,961

Notes: This table reports the mean, standard deviation, and the number of observations for each variable. Use of force and use of force with a gun are measured at the call level and take on values of one if the call ended in a use of force or use of force with a gun. Overall, 7 percent of use of force involves a gun, 38 percent involves a taser, and the remaining 55 percent is a category that includes the officer using things like his hands, feet, baton, mace, or nightstick to subdue the civilian. Priority, latitude, and longitude have been altered (multiplied by a random number) to protect the city's anonymity. Standard deviations in parentheses.

The police force reflects these demographics, as 85 and 83 percent are minority and Hispanic, respectively. Force is observed at the call level, and was used on 389 calls. White and minority officers use force at a similar rate of around 1 in 1,000, which is similar to the average use of force rate in the first city.

II. Research Design and Methodology

To estimate the difference between White and minority officers in overall use of force, we estimate the following equation:

$$UseofForce_c = \beta_0 + \beta_1 I\left(WhiteOfficer\right)_c + Beat_c + X_c + \epsilon_c. \tag{1}$$

Use of force is a binary variable equal to one when a call c ends in a use of force and zero for calls that do not involve a use of force; *WhiteOfficer* takes on a value of one when the police officer is White and zero otherwise; β_1 captures the difference in the probability of force across officer race; X_c includes control variables at the call-level. Specifically, X_c includes controls for officer gender, latitude, longitude, and time between call and dispatch, as well as fixed effects for priority of call, officer experience, day of the week, call description, call taker, and home beat, as proxied by the beat in which the officer responds to the most calls. Our baseline specification includes *Beat* fixed effects in order to isolate the as-good-as-random variation in the race of the officer assigned. However, we also report estimates that include *Beat*Year*Week*Shift* fixed effects.

To estimate whether White and Black officers scale up the use of force similarly as they go from White to Black neighborhoods, we use a difference-in-difference style approach. Formally, we estimate the following for the first city in our analysis:

$$\text{(2)} \quad UseofForce_{ic} = \beta_0 + \beta_1 \left(ProportionBlackCivilians\right)_c + Officer_i + \beta_2 I \left(WhiteOfficer * ProportionBlackCivilians\right)_{ic} + Beat_c + X_c + \epsilon_{ic},$$

where $Officer_i$ is an individual officer fixed effect, and $ProportionBlackCivilians_c$ is the proportion of Black civilians in the census block group of the call and controls for differences in the underlying risk of calls across neighborhoods of different racial compositions. The variable of interest is the interaction between $ProportionBlackCivilians_c$ and $WhiteOfficer_i$. We interpret the coefficient, β_2, as the effect of dispatching a different-race officer on use of force. As discussed earlier, this measure is necessarily a relative one. It provides an estimate of how much additional force is used by White officers on Black versus White civilians, compared to if White officers increased their use of force the same as Black officers. We also note that we do not observe—and view as unknowable—whether force *should have* been used on a given call. As a result, without imposing additional assumptions we cannot know whether a nonzero effect means White officers used too much force on Black civilians, or too little on White civilians; or whether Black officers used too much force on White civilians, or too little on Black civilians. The coefficient β_2 captures the sum of those effects.

Intuitively, β_2, is identified by comparing the rate of force used in Black versus White neighborhoods for individual officers, averaging that difference across officer race, and taking the difference. The inclusion of officer fixed effects addresses the possibility that more aggressive White officers could be assigned to beats in or closer to predominantly Black neighborhoods, while less aggressive White officers sort into White neighborhoods. As a result, difference-in-difference estimates without officer fixed effects would capture the combined effect of within-officer behavior and differential sorting of White and Black officers across neighborhoods. In this way, equation (2) ensures that any nonzero effects are caused by racial differences in individual officer behavior.

In addition, we note that estimates from both approaches are necessarily identified by officers who are dispatched to calls for service. This also includes officers who are less likely to initiate traffic stops, or who spend less time on calls, as those officers are more available for a call. These may or may not be representative of the overall police force, though we would argue they are an important set of officers in their own right.

We note that in equations (1) and (2), we identify effects using only the race of the dispatched officer(s), even though others may choose to respond to the call as well. In this way, we implement an intent-to-treat analysis. Relatedly, for the analysis of the first city we observe whether force was used by the dispatched officer, or by another officer who also responded to the same call.[10] Our main analysis assigns use of force to a dispatched officer if force was used by any officer on that call. However, in online Appendix Table A6 we show estimates are almost identical when use of force is defined as force used by that dispatching officer. We also note that for some calls, multiple officers are dispatched to the call. This is relatively rare in the first city (two percent of calls), though is more common in the second city (56 percent). As a result, we include each officer-call as an observation, and weight each observation by the inverse of the number of officers dispatched to the call.

For both approaches, we report standard errors that are clustered at the officer level to allow observations to be correlated across cases for a particular officer.[11] In addition, we report [in square brackets] standard errors that are block bootstrapped at the officer level. Finally, in the text we also report empirical p-values from randomization inference where we randomly assign race to each officer in our sample.[12] We note, however, that while randomization inference is often favored by practitioners, econometricians and statisticians have pointed out problems with its application to nonexperimental settings (e.g., Efron 2004). This is due to the inability to replicate the process that generated the observed data. This applies to our setting, as there is no known procedure for assigning officers to neighborhoods, which means the data we observe are one realization of an unknown data generating process.[13] As a result, we prefer inference using the clustering or resampling techniques proposed by Bertrand, Duflo, and Mullainathan (2004).

We estimate effects using ordinary least squares (OLS), though we also report odds ratios from a logit model as a robustness exercise. We prefer OLS for two reasons. First, in order to attain convergence of the logit model, we are only able to control for police area and month fixed effects, or police area by year fixed effects, rather than beat-by-year-by-week-by-shift fixed effects. Police areas are roughly five times as big as a beat in the first city, and more than ten times as large in the second city. Perhaps more problematically, given a fixed number of observations in each group, the logit estimator is not consistent as the number of groups increases

[10] For the second city, we only observe whether the dispatched officer used force.

[11] A previous version of the paper also reported standard errors that were two-way clustered at the officer and beat levels, which were very similar.

[12] We randomly assign race to individual police officers. Then we then estimate and record the coefficient of interest. We repeat this exercise for 1,000 iterations and report the empirical p-value, which is the fraction of t-statistics from randomized estimates more extreme (i.e., larger in absolute value) than our estimate from the real data.

[13] This contrasts with other settings such as jury panel assignment studied by Hoekstra and Street (2021), who can replicate the known random assignment procedure and thus estimate the distribution of possible coefficients.

(Chamberlain 1980, Heckman 1987). This is of particular concern for specifications that include officer fixed effects.

Intuitively, the difference-in-difference approach outlined in equation (2) compares differences in the probability of use of force between Black civilians and White civilians for Black officers and White officers. Importantly, this model allows for encounters with Black civilians to merit more or less force than interactions with White civilians. In this way, our approach differs significantly from approaches that rely on controlling for observable contextual factors to account for selection in the type of interactions with White versus Black civilians. Rather, we identify different-race effects by comparing how a given White officer's use of force differs when he is dispatched to a more Black versus less Black neighborhood, compared to what happens when a given Black officer is dispatched to more and less Black neighborhoods. In short, we include beat or beat-by-time effects to ensure White and Black officers are dispatched to similar calls, and individual officer fixed effects to ensure differences in use of force across civilian race are not driven by nonrandom sorting of officers across neighborhoods.

As alluded to earlier, our coefficient of interest β_2 captures the net impact of several potential mechanisms. These include racial bias by police officers, such as applying a lower threshold for the use of force in different-race neighborhoods. However, the effect could also capture differential skill, such as officers being better at correctly interpreting behavior in same-race versus different-race neighborhoods. For example, White officers may misperceive the safety or threat in Black neighborhoods. This is consistent with experiments by Barsamian Kahn and Davies (2017), who find that error rates in the decision to "shoot" Black civilians are lower in perceived safe neighborhoods. Importantly, we note that racial bias and differential skill by officers could induce different civilian responses that could itself make force more or less likely, though we attribute that to the officer rather than to the civilian. However, β_2 also captures the net impact of civilian response to officer race per se. For example, it is possible that civilians would respond differently to different-race officers *even if those officers behaved in exactly the same way*. If civilians respond more aggressively to different-race officers, ceteris paribus, that behavior could explain some or all of the estimate. On the other hand, civilians may be more cooperative with different-race officers, in which case controlling for that behavior would increase our estimate. We view both responses as possible, and remain agnostic about the impact of either on the estimates we report.[14]

Both the cross-sectional approach and the difference-in-difference approach rely on the assumption that conditional on a beat or beat-by-time fixed effect, the variation in police officer race is as good as random. Given our understanding of dispatch protocol in both cities, we believe there are ex ante reasons to believe this assumption is valid. In addition, we also empirically assess the validity of our research design.

[14] We are skeptical regarding the extent to which researchers are capable of distinguishing the impact of police behavior from civilian behavior since speech, tone, and body language are typically not observed in the data. To our knowledge the best evidence to date comes from Voigt et al. (2017), who analyze officer speech from officers' body cam footage in Oakland during traffic stops. They report that officers speak more disrespectfully to Black civilians than White ones, controlling for observed contextual factors. The authors note it is possible for at least some of this effect to be caused by civilian behavior, though they argue it is unlikely all due to civilian behavior given the difference is present in the first 5 percent of words spoken by the officers, and is judged to be present even when evaluated in the context of what was said by the civilian.

Table 2—Correlation between Call Characteristics and Officer Race

	Proportion Black civilians (1)	Per capita income (2)	Proportion unemployed (3)	Proportion less than high school degree (4)	Call priority (5)	Time between call and dispatch (6)	Call from home beat (7)
Panel A. Unconditional							
White officer	−0.0279	788.8	−0.00598	−0.00346	−0.000808	−0.0463	−0.00546
	(0.0185)	(447.9)	(0.00394)	(0.00309)	(0.0122)	(0.143)	(0.00502)
Observations	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139
Outcome mean	0.586	23,281.7	0.139	0.185	2.839	6.490	0.180
Panel B. Beat fixed effects							
White officer	−0.00260	110.3	−0.000582	−0.000217	−0.00468	0.0694	−0.00711
	(0.00183)	(107.8)	(0.000399)	(0.000732)	(0.0106)	(0.117)	(0.00430)
Observations	1,233,139	12,33139	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139
Outcome mean	0.586	23,281.7	0.139	0.185	2.839	6.490	0.180
Panel C. Beat-year-week-shift fixed effects							
White officer	−0.00128	52.41	−0.000247	−0.000208	−0.0105	−0.108	−0.00582
	(0.000662)	(43.58)	(0.000246)	(0.000254)	(0.00857)	(0.144)	(0.00374)
Observations	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139	1233139
Outcome mean	0.586	23,281.7	0.139	0.185	2.839	6.490	0.180

Notes: This table reports the coefficient on *WhiteOfficer* from separate regressions of call characteristics on a binary variable representing officer race. Panel B includes beat fixed effects, and panel C includes beat-year-week-shift fixed effects. Standard errors are clustered at the officer level. Priority has been altered (multiplied by a random number) to protect our city's identity. Standard errors in parentheses.

First, when we estimate effects, we will examine the extent to which adding controls affects our estimates of interest. Specifically, we control for call characteristics including call priority, latitude, longitude, time between call and dispatch, as well as fixed effects for day of the week, call description, call taker, and home beat. We also add controls for officer gender and years of experience. If our identifying assumption is valid, we expect adding these controls should not affect our coefficient of interest.

The second way in which we assess the validity of our research design is to examine directly the correlation between call characteristics and officer race. Specifically, we test whether officer race is correlated with the racial composition of the neighborhood, the call priority, time between call and dispatch, whether the call came from an officer's home beat, and other census block group characteristics (i.e., per capita income, proportion unemployed, and proportion with less than a high school degree) (United States Census Bureau 2016a, b, c). We formally test this in Table 2, where we regress each of these characteristics on an indicator for whether the dispatched officer was White. In panel A, we report estimates from specifications that include no other controls. In panel B we additionally control for beat fixed effects, and in panel C we control for beat-by-year-by-week-by-shift fixed effects. Panel A shows that there is relatively little sorting by police officer race in this city. In particular, of the seven coefficients, only one is significant at the 10 percent level (per capita income, coefficient = \$789) and none are significant at the five percent level. Similarly, panels B and C show little evidence of correlation between officer race and call characteristics once conditioning on beat and beat-by-time fixed effects, respectively. In both cases only one of seven coefficients is significant at the 10 percent level (proportion Black civilians). In addition, the economic magnitude of the coefficients is small. For example, the marginally significant estimate in panel C indicates that White officers are dispatched to calls in areas that are 0.1 percentage points less Black than the calls to which Black officers are dispatched. Collectively, the lack of statistical and economic significance of the coefficients reported in Table 2 is consistent with the identifying assumption of our study.[15]

In addition, the third way in which we assess our research design is to use all call characteristics to predict officer use of force. Specifically, we first regress police use of force on beat-year-week-shift fixed effects. We then regress these residuals—which capture the deviation from the average use of force for that beat and time—on every covariate we observe for each call. These include proportion Black civilians in the block group, call priority, latitude, longitude, time between call and dispatch, home beat, per capita income, proportion of civilians with less than high school degree and proportion unemployed, as well as fixed effects for call description and call taker. We use the resulting regression equation to predict the likelihood force would be used for each officer on each call. We then add the mean use of force rate for the full sample to each predicted value. Intuitively, this produces a linear combination of exogenous call characteristics, where the weights are chosen as to best predict the likelihood of force being used. We then ask whether White and Black officers are dispatched to calls of similar underlying danger when assigned to a neighborhood of a given racial composition. If the identifying assumption of our approach is valid, predicted use of force should be the same for White officers as Black officers.

We show results of this test graphically for the first (Black/White) city in Figure 2. Panel A shows results for all use of force, while panel B shows results for only use of force with a gun. In both cases, results demonstrate that conditional on the police beat-year-month-week-shift of the call, White and Black officers are dispatched to calls that are of similar underlying risk.[16] This is consistent with the identifying assumption, and with our understanding of how officers are dispatched.

Finally, for the difference-in-difference approach, we also add controls for interactions between officer race and all call characteristics. We do so in order to shed light on the mechanism underlying the different-race effects. In particular, we test whether the effect can be explained by officers having an increased propensity to use force for the type of calls that occur in different-race neighborhoods. For example, if White officers were more likely to use force when dispatched to domestic disturbance incidents, and if domestic disturbance incidents were more likely to occur in Black neighborhoods, that could generate a nonzero difference-in-difference estimate. It is important to note, however, that because this behavior has a disparate impact on different-race civilians, the effect is the same as explicit bias.

[15] Similarly, in online Appendix Table A5 we show results from a difference-in-difference specification where we regress call and neighborhood characteristics on an indicator for White officers, proportion Black civilians, and the interaction. The table, which follows the same form as Table 2, shows that none of the 24 coefficients on the interaction term are significant at even the 10 percent level.

[16] We also note that the neighborhood and call characteristics used to predict use of force explain around 1–1.5 percent of the variation in use of force and use of force with a gun.

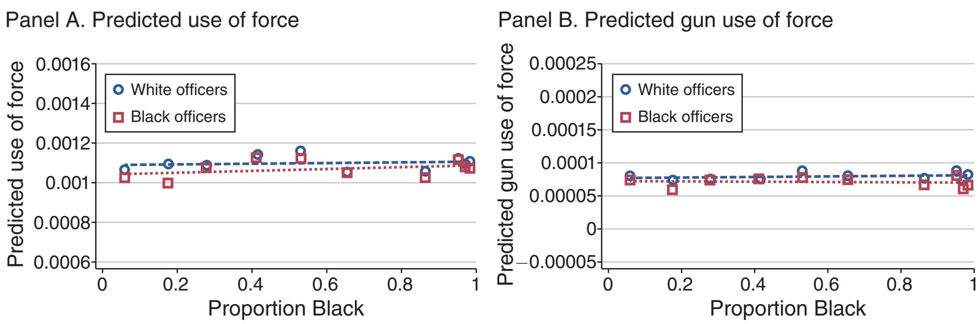

Figure 2. Predicted Outcomes for Black and White Officers

Notes: In panels A and B, we predict the probability of use of force or use of force with a gun using all observable call characteristics for each call for service. Specifically, we use proportion Black civilians to predict (after removing beat-year-week-shift fixed effects) unemployment, per capita income, proportion high school dropouts, latitude, longitude, and time between call and dispatch; as well as fixed effects for the day of the week, priority, call description, home beat, and call taker using a linear probability model. Observations are grouped so that each point includes an equal number of calls. The fitted line is a linear fit across all predicted use of force rates.

III. Results

We begin by showing results for the first city graphically. Results for all use of force are shown in Figure 3. Each graph shows local averages of use of force by race of officer, as represented by the blue circles (White officers) and red squares (Black officers). Each circle/square includes the same number of calls. We also fit lines to the underlying data, by officer race.

Panel A of Figure 3 shows actual use of force by officer race. It reveals two main takeaways. The first is that regardless of the racial composition of the neighborhood, White officers are more likely to use force than Black officers. This suggests that with respect to the likelihood of using force, White officers seem to be drawn from a different distribution than Black officers. The second takeaway from panel A of Figure 3 is that while the propensity of Black officers to use force increases only modestly as they are dispatched to neighborhoods with higher proportions of Black civilians, White officers use significantly more force as they are dispatched to more Black neighborhoods. This suggests that dispatching an different-race officer to a scene increases the likelihood that force will be used. For example, if officer race (White versus Black) mattered for use of force, but having an officer of a different race did not, we would expect parallel slopes for White and Black officers in Figure 3. The difference in slopes suggests that dispatching an different-race officer to a call results in a higher likelihood of police use of force. Panel A of Figure 3 shows that while White officers use force approximately 25 percent more in all-White neighborhoods, they use force around 85 percent more in all-Black neighborhoods.

Panel B of Figure 3 shows actual use of force with a gun by officer race. It shows that while Black and White officers have roughly similar propensities to fire their guns when assigned to majority-White neighborhoods, they differ significantly when dispatched to neighborhoods with at least 80 percent Black residents. In those neighborhoods, White officers are roughly five times more likely to use gun force compared to Black officers. Again, this suggests that White officers are more

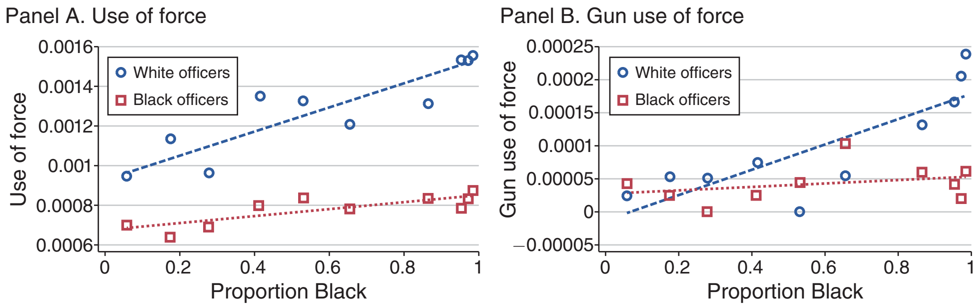

Figure 3. Actual Use of Force for Black and White Officers

Notes: In panel A we plot use of force. In panel B we plot gun use of force. The fitted line is a linear fit across all use of force rates. Observations are grouped so that each point includes an equal number of calls.

likely to use force overall and that they are especially more likely to use force in mostly-Black neighborhoods, suggesting large different-race effects.[17]

We note that one potential downside of the unconditional means in Figure 3 is that there is the potential for nonrandom sorting within neighborhoods of similar race to generate bias. For example, if White officers were systematically assigned to the neighborhoods within a given racial composition that generate the most dangerous calls, that could bias estimates of how Black and White officers differ in their propensity to use force. We note that results in panel A of Table 2 and online Appendix Figure A1 indicate there is little nonrandom sorting of officers on observable characteristics, which suggests even an unconditional analysis may well be valid. Nevertheless, we now turn to estimating differences by officer race controlling for beat and beat-by-time fixed effects, and estimating difference-in-difference specifications that include those fixed effects as well as individual officer fixed effects.

A. *Main Results*

We first estimate the difference in the propensity of White and Black officers to use force. Estimates of the effect of exposure to a White officer compared to a Black officer are shown in columns 1 and 2 of Table 3, where each column represents a different regression. Panel A shows results for all use of force, while panel B shows results for use of force with a gun. Column 1 reports results from a regression of use of force on an indicator for White officer, along with beat fixed effects. Column 2 controls for beat-by-year-by-week-by-shift fixed effects and call characteristics including time between call and dispatch, per capita income, unemployment, and proportion with less than a high school degree, as well as fixed effects for day of the week, priority of call, call description, call taker, officer gender, officer years of experience, and officer home beat, as proxied by the beat to which he responded to the most calls.

[17] Online Appendix Tables A1 and A2 show the underlying number of incidents of force with a gun that along with number of calls generate the rates shown in panel B of Figure 3. These numbers are shown for calls in neighborhoods that are less than 25 percent Black, 25–75 percent Black, and more than 75 percent Black.

Table 3—The Effect of Officer Race and Different-Race Officers

	Officer race				Difference-in-differences Different race officer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Use of force						
White officer	0.000485	0.000428	0.000256	0.000116		
	(0.0000796)	(0.0000791)	(0.000122)	(0.000130)		
	[0.0000831]	[0.0000853]	[0.000126]	[0.000139]		
Proportion Black civilians					−0.000102	−0.000368
					(0.000230)	(0.000325)
					[0.000240]	[0.000336]
White officer × proportion Black civilians			0.000387	0.000525	0.000536	0.000629
			(0.000169)	(0.000188)	(0.000270)	(0.000275)
			[0.000171]	[0.000193]	[0.000280]	[0.000291]
Observations	1,233,139	1,233,139	1,233,139	1,233139	1,233,139	1,233,139
Outcome mean	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106
Panel B. Gun use of force						
White officer	0.0000540	0.0000460	−0.0000298	−0.0000551		
	(0.0000219)	(0.0000228)	(0.0000222)	(0.0000245)		
	[0.0000231]	[0.0000252]	[0.0000255]	[0.0000298]		
Proportion Black civilians					−0.000162	−0.000116
					(0.0000961)	(0.0000787)
					[0.000106]	[0.0000846]
White officer × proportion Black civilians			0.000142	0.000170	0.000357	0.000368
			(0.0000472)	(0.0000530)	(0.000140)	(0.000133)
			[0.0000516]	[0.0000596]	[0.000150]	[0.000143]
Observations	1,233,139	1,233,139	1,233,139	1,233,139	1,233,139	1233,139
Outcome mean	0.0000710	0.0000710	0.0000710	0.0000710	0.0000710	0.0000710
Beat FE	Yes	—	Yes	–	Yes	–
Beat-year-week-shift FE	—	Yes	—	Yes	–	Yes
Call controls	—	Yes	–	Yes	–	Yes
Officer FE	—	—	–	–	Yes	Yes

Notes: This table shows the effect of officer race (columns 1–4) and different-race officers (columns 5 and 6) on use of force (panel A) and gun use of force (panel B). Even columns add controls for the time between call and dispatch, latitude, longitude, per capita income, unemployment, and proportion with less than a high school degree, as well as fixed effects (FE) for the day of the week, the priority of the call, call description, call taker, officer gender, officer years of experience, and officer home beat, as proxied by the beat to which the officer responded to the most calls. In columns 5 and 6, individual officer fixed effects subsume *WhiteOfficer*. Standard errors clustered at the officer level are reported in parentheses, and cluster bootstrap standard errors are in brackets.

Results in column 1 of panel A indicate that White officers are 0.0485 percentage points more likely to use force than Black officers. Given the average use of force of 0.106 percent, this suggests that White officers are 46 percent more likely to use force relative to the mean, and 64 percent more likely relative to the mean for Black officers of 0.076 percent, as shown in Table 1. The estimate in column 2 changes only slightly to 0.0428 percentage points, which represents a 56 percent increase relative to the mean for Black officers. Both estimates are significant at the 1 percent level using standard errors that are clustered at the officer level or block bootstrapped. Estimates in panel B show even larger differences for use of force with a gun; estimates in columns 1 and 2 indicate White officers are 136 and 116 percent more likely to fire their guns than the mean for Black officers. The estimate in column 1 is significant at the 5 percent level using either clustered or block bootstrapped standard errors, while the estimate in column 2 is significant at the 5 (10) percent level using clustered (bootstrapped) standard errors.

Next, we measure the extent to which the increased average propensity to use force by White officers is driven by increased force in Black neighborhoods, rather than White ones. We do so by regressing an indicator for use of force on an indicator for White officer and an indicator for White officer times proportion of citizens in that block group who are Black. Results are shown in columns 3 and 4 of Table 3, where column 3 controls only for beat fixed effects and column 4 controls for both beat-by-time fixed effects and call characteristics. Results provide strong evidence that effects are driven by increased force in Black neighborhoods, consistent with Figure 3. Estimates in panel A indicate that White officers are 0.0387 and 0.0525 percentage points more likely to use force in Black neighborhoods, which are significant at the 5 and 1 percent levels, respectively. In contrast, the difference in White neighborhoods is much smaller at 0.0256 and 0.0116 percentage points, respectively. Again, both estimates are significant at the 1 percent level. Estimates in panel B show that the positive estimates in the first two columns for use of force with a gun are also driven entirely by increased use of force in Black, rather than White neighborhoods. Estimates on the interaction term are 0.014 and 0.017 percentage points, both of which are significant at the 1 percent level using either clustered or bootstrapped standard errors and are economically large relative to the baseline rate of use of force by Black officers of 0.00397 percent.

Finally, we estimate the difference-in-difference specification in columns 5 and 6. Column 5 controls only for beat fixed effects, while column 6 controls for both beat-by-time fixed effects and call characteristics. Importantly, both columns control for officer fixed effects. In this way, estimates capture only the effect of individual officers being more likely to use force in different-race neighborhoods, rather than the joint effect of both that and any nonrandom sorting of White and Black officers across neighborhoods. Estimates in columns 5 and 6 of panel A are 0.0536 and 0.0629 percentage points, which indicate different-race effects of 51 and 59 percent, respectively. The former is significant at the 5 and 10 percent level using clustered and bootstrapped standard errors, while the latter is significant at the five percent level using either. Effects are even larger for use of force with a gun; estimates in panel B indicate that White officers are scaling up use of force with a gun 5.0–5.2 times more than Black officers. Estimates are significant at the 5 and 1 percent levels using clustered standard errors, respectively, and each is significant at the 5 percent level using bootstrap.[18]

In summary, results in columns 1–4 show that White officers use any force and force with a gun 60 and 100 percent more than Black officers, and that these differences are driven by increased use of force in predominantly Black neighborhoods. In addition, results in columns 5 and 6 indicate that individual officers are 55 percent more likely to use any force and five times more likely to use force with a gun in different-race neighborhoods.

[18] We also performed randomization inference, though we prefer clustering or resampling for reasons discussed earlier. We compute two-sided p-values of 0.000 and 0.000 for the estimates in columns 1 and 2 of panel A, and 0.008 and 0.037 for the same columns in panel B. Similarly, for the interaction terms reported in columns 3–6 we calculate empirical two-sided p-values of 0.022, 0.006, 0.052, and 0.025 in panel A and 0.001, 0.000, 0.001, and 0.001 in panel B.

B. *Robustness*

Next, we examine the robustness of the results in Table 4. One potential concern with the outcome of police use of force is that there could be a reporting bias, especially for the type of force that is on the margin of qualifying as use of force. For example, if officers are more likely to report otherwise-similar use of force when it is used in different-race neighborhoods, that could generate the nonzero difference-in-difference estimates in Table 3. To address this concern, in panel C of Table 3 we show results for use of force from either a taser or a firearm. We focus on these types of force because they are the types where misreporting would be most difficult, if not impossible. In both cases, there is ex post objective evidence on whether that action was taken. For reference, we replicate our main results for use of force and use of force with a gun in columns 1, 3, and 5 of panels A and B. Estimates in panel C are similar in both significance and relative magnitude. We estimate that White officers are 45 percent more likely to use a taser or a gun compared to Black officers, which is significant at the one percent level. It is also of similar magnitude as for all use of force as reported in panel A, which was 64 percent. Column 3 of panel C shows that as for results for all use of force and use of force with a gun, this increased force is driven primarily by additional force being used in Black neighborhoods. Column 5 shows the difference-in-difference estimate for use of force with a taser or gun and indicates that different-race officers use force 124 percent more relative to the mean, which is significant at the one percent level. Collectively, estimates in columns 1, 3, and 5 of column C indicate the main results for the first city are unlikely to be due to the systematic misreporting of use of force.

Next, we report odds ratios from logit estimation. Corresponding specifications are shown in columns 2, 4, and 6 of Table 4. Estimates of the average difference between White and Black officers are similar to those from OLS. Odds ratios in column 2 are all significant at the one percent level and imply 65 and 138 percent increases for all force and use of force with a gun, relative to Black officers. Odds ratios in column 4 show that while the difference for use of force with a gun is due to increased force in Black neighborhoods (odds ratio = 4.86), it is less clear for use of force overall where the odds ratios are similar for White and Black neighborhoods. Similarly, while the odds ratio for the difference-in-difference estimate suggests a similar magnitude effect for all use of force (odds ratio = 1.54), it is not significant at conventional levels. However, odds ratios for the difference-in-difference estimate from the officer fixed effect specification is even larger (17.67) for gun use of force and for taser or gun use of force (2.55), both of which are significant at the 5 percent level. Overall, we conclude that the results we document using logit estimation are similar in magnitude to those from OLS reported in Table 3, which we prefer in light of the incidental parameters problem with the logit estimator in settings like this (Chamberlain 1980, Heckman 1987).

We also assess the degree to which our difference-in-difference estimates could be driven by White officers scaling up use of force based on some officer or call characteristic correlated with race, rather than race itself. Similarly, we ask whether estimates are driven by less experienced or male officers—both of which are correlated with officer race, as shown in Table 1—scaling up force more as they go to more Black neighborhoods. To do so, we add interactions between each call characteristic

Table 4—Robustness

	Officer and civilian race				Difference-in-differences Different race officer		
	(1)	Logit (2)	(3)	Logit (4)	(5)	Logit (6)	(7)
Panel A. Use of force							
White officer	0.000485	1.646	0.000256	1.390			
	(0.0000796)	(0.136)	(0.000122)	(0.188)			
Proportion Black civilians					−0.000102	0.946	3.935
					(0.000230)	(0.251)	(36,039.1)
White officer × proportion Black civilians			0.000387	1.306	0.000536	1.541	0.000560
			(0.000169)	(0.214)	(0.000270)	(0.440)	(0.000314)
Observations	1,233,139	1,223,413	1,233,139	1,223,413	1,233,139	829,428	1,233,139
Outcome mean	0.00106	0.00110	0.00106	0.00110	0.00106	0.00162	0.00106
Panel B. Gun use of force							
White officer	0.0000540	2.381	−0.0000298	0.737			
	(0.0000219)	(0.765)	(0.0000222)	(0.498)			
Proportion Black civilians					−0.000162	0.483	1.594
					(0.0000961)	(0.569)	(21,271.6)
White officer × proportion Black civilians			0.000142	4.861	0.000357	17.67	0.000332
			(0.0000472)	(3.763)	(0.000140)	(24.19)	(0.000132)
Observations	1,233,139	739,113	1,233,139	739,113	1,233,139	71,418	1,233,139
Outcome mean	0.0000710	0.000127	0.0000710	0.000127	0.0000710	0.00132	0.0000710
Panel C. Reportable use of force							
White officer	0.000212	1.615	−0.00000125	1.066			
	(0.0000473)	(0.174)	(0.0000657)	(0.197)			
Proportion Black civilians					−0.0000612	1.138	2.262
					(0.000173)	(0.489)	(29125.4)
White officer × proportion Black civilians			0.000361	1.862	0.000585	2.548	0.000471
			(0.000110)	(0.446)	(0.000211)	(1.180)	(0.000231)
Observations	1,233,139	1,190,747	1,233,139	1,190747	1,233,139	567,568	1,233,139
Outcome mean	0.000470	0.000501	0.000470	0.000501	0.000470	0.00105	0.000470
Beat fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Call controls	—	—	—	—	—	—	Yes
Officer fixed effects	—	—	—	—	Yes	Yes	Yes
Interactions	—	—	—	—		—	Yes

Notes: This table shows the effect of officer race (columns 1–4) and different-race officers (columns 5–7) on use of force (panel A), gun use of force (panel B), and either gun or taser use of force (panel C). Even columns (logit specifications) report odds ratios. In columns 5, 6, and 7, individual officer fixed effects subsume *WhiteOfficer*. Standard errors are clustered at the officer level. Column 7 add controls for the time between call and dispatch, latitude, longitude, per capita income, unemployment, and proportion with less than a high school degree; as well as fixed effects for the day of the week, the priority of the call, call description, call taker, officer gender, officer years of experience, and officer home beat, as proxied by the beat to which the officer responded to the most calls. Interactions between call characteristics and officer race as well as interactions between proportion Black civilian and officer gender and years of experience are also added. In logit specifications, beats (columns 2, 4) and officers (column 6) with no use of force are dropped.

and officer race, as well as between proportion Black and officer gender, and proportion Black and years of experience. Results are shown in column 7. Estimates are similar for all use of force (0.056 versus 0.054 percentage points) and for use of force with a gun (0.033 versus 0.036 percentage points).[19] This suggests our findings are not driven by officers scaling up force on the basis of an observed call

[19] We note that because we are also including interactions between proportion Black civilian and other variables, the coefficient on proportion Black civilians is not easily interpreted.

characteristic, like priority level or type of call, rather than civilian race. Similarly, it demonstrates that estimates are not driven by any differential scaling up of force by male officers or by less experienced officers.

We also examine the extent to which difference-in-difference estimates on use of force with a gun are sensitive to any individual officer. In online Appendix Figure A6, we report the coefficients that result from dropping any one of the officers involved in the 94 incidents of force with a gun observed in the data. We do so for specifications shown in columns 1 through 6 of panel B in online Appendix Table A6. Of the six estimates of interest, the only one that drops out of significance is the marginally significant coefficient of 0.00460 percentage points shown in column 2, which is the average difference between White and Black officer use of force with a gun when controlling for beat-by-time fixed effects and call characteristics. In contrast, estimates showing that this increase is driven by additional force in Black neighborhoods and the within-officer difference-in-difference estimates remain significant and of similar magnitude when any given officer is dropped.

C. *Individual Officer Effects*

Given the findings discussed above, it is natural to ask whether the differences by race are due to only a handful of officers, or if they are more systemic. Put differently, are the effects we find due to differences in the middle of the distribution, or are they due to differences in the tails? To address this question, we estimate an individual officer random effects model, and then compute and graph the distribution of (shrunken) effects for White and Black officers. We begin by regressing use of force on beat-by-time fixed effects and controls as in column 4 of Table 3, and keep the residuals. We then use those residuals and the Stata command *mixed* to estimate a random effects model, and then compute the random effect for each officer. We do so only for officers who respond to at least 500 calls, which limits the sample to 46 percent of officers in our sample, though those officers respond to 91 percent of the calls. After estimating officer effects, we trim the 5 percent most extreme officers in the far tails so the reader can visualize any differences in the distributions.[20]

Results comparing Black and White officers are shown in Figure 4, panel A, which shows a kernel density plot of both White and Black officers. Figure 4, panel A shows a rightward shift in the distribution of White officers, compared to Black officers. This suggests that the increased propensity to use force by White officers is not driven by a handful of officers, but is rather due to an increased propensity across a significant fraction of White officers.

Figure 4, panel B shows that in majority-Black neighborhoods, there are substantially fewer White officers on the left side of the distribution and more in the middle and right tail of the distribution, compared to Black officers. Figure 4, panel C shows that in White neighborhoods, there is a shift in the White officer distribution from the middle of the distribution toward the right tail.

We can also use the individual officer estimates shown in Figure 4 to address several other questions. First, we attempt to quantify the proportion of officers who

[20] Without trimming, the resulting figures are so "zoomed out" that it is impossible to distinguish between the distributions even when there are meaningful differences.

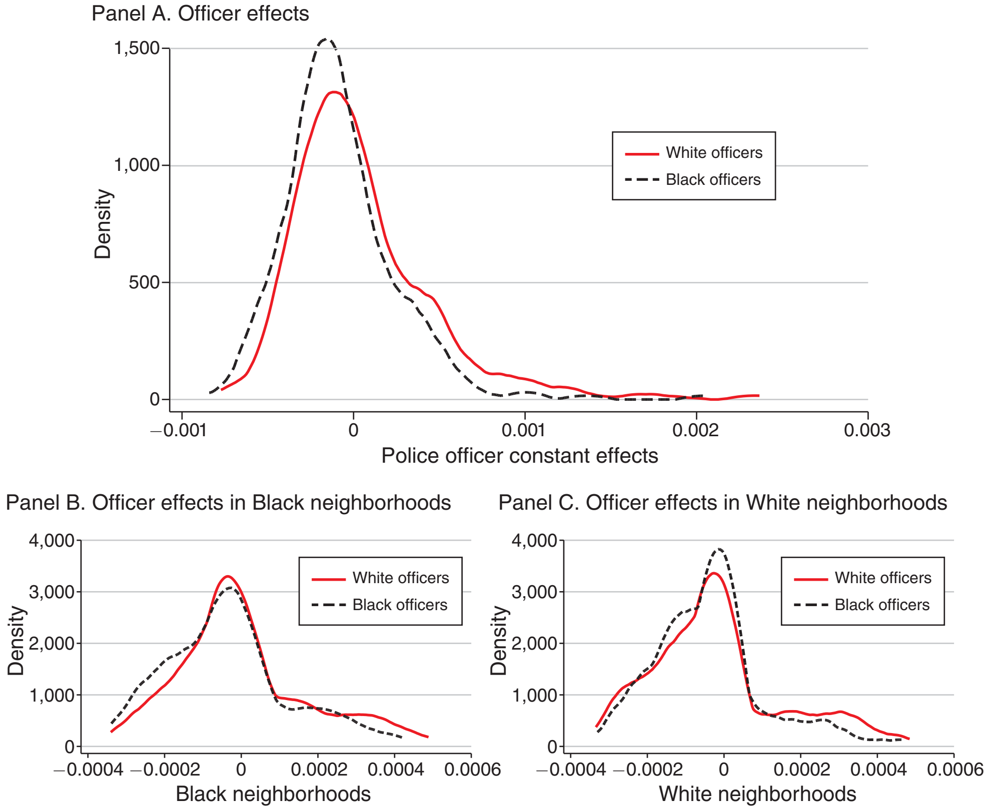

Figure 4. Individual Officer Effects

Notes: Figures represent the distribution of individual police officer effects (random effects) by police officer race (panel A) and neighborhood race (panel B) and (panel C). Random effects are calculated by first removing the controls and fixed effects added in even columns in Table 3. Specifically we include controls for the time between call and dispatch, latitude, longitude, per capita income, unemployment, and proportion with less than a high school degree; as well as fixed effects for the day of the week, the priority of the call, call description, call taker, officer gender, officer years of experience, and officer home beat, as proxied by the beat to which the officer responded to the most calls. Then a constant random effect is calculated (using residualized use of force) for each officer in panel A or a constant for each officer and neighborhood they respond to. Each panel shows results from a different regression and only officers responding to more than 500 calls.

are driving the overall effects documented in Table 3. To do so, we sequentially drop White officers from the full sample, beginning with the officers in the right tail of the distribution in Figure 4, until the overall estimate goes to zero. Thus, in order to assess how many officers are driving the result that White officers use force 50–60 percent more than Black officers on average, we begin dropping White officers from the right tail of Figure 4, panel A. Results indicate that we need to eliminate 32 percent of White officers (150 out of 474 officers for whom we computed a random effect) in order to reduce the propensity of using force for White officers to that of Black officers (i.e., to reduce the estimate in column 1 in panel A of Table 3 to zero). Similarly, results indicate we would need to eliminate 31 percent of White officers (145 out of 474) in the right tail of Figure 4 panel B in order to reduce the different-race effect we document in column 5 of Table 3 to zero. This suggests that the aggregate effects we pick up in our main analysis are not driven by a small fraction of officers.

In addition, we can also use the results in Figure 4 to identify whether there are more officers in the tails of the distribution of overall use of force, or use of force in Black neighborhoods, than we would expect even given the number of officers in the sample. To do so, we replicate the method proposed by Efron (2004), as implemented by Ridgeway and MacDonald (2009) in identifying police officers in the tails of the distribution in terms of making pedestrian stops. This method involves first fitting a normal distribution over the center mass of the observed distribution of officer effects, as well as estimating the actual density function.[21] We note that in doing so, this approach necessarily assumes that officers labeled by Efron (2004) as "interesting" are in the tails of the distribution, rather than the center mass. For each officer in the distribution, we then compute a false discovery rate (fdr), which is defined as the null/actual density. The odds of being problematic are then defined as 1–fdr. Following Ridgeway and MacDonald (2009), we flag officers who have false discovery rates of less than 0.5. These are the officers who appear to be using force at a substantially higher rate than their statistical benchmark.

We note that our setting has a significant advantage over most settings in which researchers attempt to identify the impact of individual agents. This includes Ridgeway and MacDonald (2009), though it also includes research in other contexts such as teacher quality (Chetty, Friedman, and Rockoff 2014). This is because we examine two settings in which there is an institutional arrangement that assigns officers to situations in a conditionally random way, which eliminates the primary source of bias in estimating individual effects. Additionally, we note that the method outlined here, combined with a dispatch system that removes discretion such as those used in our two cities, could be used by nearly any large city to provide a straightforward way of assessing individual officers.

Results are shown for the full set of officers (Black and White) in Figure 5. We estimate that of the 748 officers who were dispatched to at least 500 calls, 75 (or 10.0 percent) were sufficiently far out in the tail as to be identified as "interesting" using the Efron (2004) method (i.e., as having a false discovery rate of less than 0.5), as implemented by Ridgeway and MacDonald (2009). All of these officers were in the right tail of the distribution, suggesting that they used force more often than one might expect even given the number of officers being tested. Of those 75 officers, 60 are White (composing 13 percent of the 474 White officers), and 15 were Black (composing 5 percent of the 274 Black officers). In addition, we can also perform the same exercise for White and Black officers separately, in both White and Black neighborhoods. These results are shown in panels B through E of Table 5. Of particular interest is the analysis in Black neighborhoods, where we flag 14 percent of White officers (67 out of 474) and only 4 percent of Black officers (11 out of 274) as using too much force relative to the null distribution. We note that this is a much higher fraction of total officers than in Ridgeway and MacDonald (2009), who flagged only 15 of 2,756 New York City officers (0.5 percent) for stopping too many (13) or too few (2) non-White pedestrians. Finally, we note that this procedure

[21] We do so using central matching (the null distribution) and a local polynomial (the actual density), as implemented by the command locfdr in R.

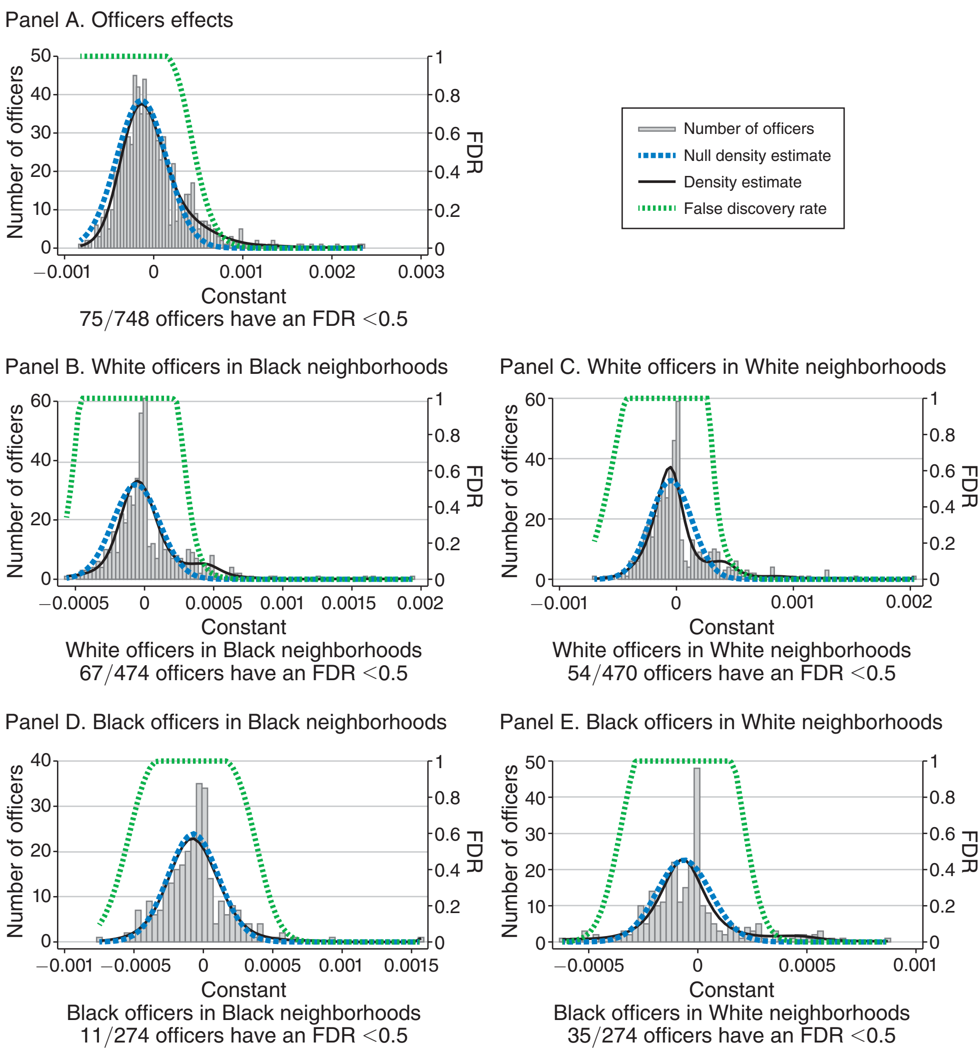

Figure 5. False Discovery Rate Individual Officers

Notes: Each figure plots a histogram of the officers effects in Figure 4 and the null density estimate used to calculate false discovery rates (FDR). Local false discovery rates were calculated following Efron (2004), and the *locfdr* package.

flags roughly equal proportions of Black and White officers in the right tail of the use of force distribution in White neighborhoods, at 9 and 11 percent, respectively.

In summary, our analyses of individual officer effects in Figures 4 and 5 provide two main results. First, Figure 4 shows that the increased propensity of White officers to use force and to scale up use of force as they go to Black neighborhoods is due to increased use of force by a sizable fraction of White officers, rather than a handful. Similarly, Figure 5 indicates that consistent with that finding, implementing an approach to flag "interesting" officers in the tails of the distribution results in the identification of roughly 10 percent of total officers, who are disproportionately White. This is driven largely by a relatively high share of White officers (14 percent) in the far right tail of the use of force distribution in Black neighborhoods.

Table 5—**Second City** The Effect of Officer Race and Different-Race Officers

	Officer race				Difference-in-differences Different race officer	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Use of force						
White officer	0.0000774	0.0000730	−0.000682	−0.000572		
	(0.0000831)	(0.0000783)	(0.000313)	(0.000313)		
Proportion minority civilian					−0.000761	−0.000562
					(0.000656)	(0.000784)
White officer × proportion minority civilians			0.000912	0.000775	0.000925	0.000716
			(0.000380)	(0.000385)	(0.000386)	(0.000392)
Observations	938,562	938,562	938,562	938,562	938,562	938,562
Outcome mean	0.000940	0.000940	0.000940	0.000940	0.000940	0.000940
Beat fixed effects	Yes	—	Yes	—	Yes	—
Beat-year-month, beat-shift fixed effects	—	Yes	—	Yes	—	Yes
Call controls	—	Yes	—	Yes	—	Yes
Officer fixed effects	—	—	—	—	Yes	Yes

Notes: This table shows the effect of officer race (columns 1–4) and different-race officers (columns 5 and 6) on use of force. Even columns add controls for latitude, longitude, per capita income, unemployment, and proportion with less than a high school degree, as well as fixed effects for the day of the week, hour dispatched, call description, call source, multi-agency, officer years of experience, officer gender, and officer home beat, as proxied by the beat to which the officer responded to the most calls. Standard errors clustered at the officer level are reported in parentheses.

D. *Results from the Second City: Whites and Hispanics*

In addition to studying the effects of police officer race in a city in which citizens and officers are mostly White or Black, we also study it in the context of a second city composed primarily of Whites and Hispanics. Figure 6 shows the distribution of calls across neighborhoods of differing race. As shown there, the vast majority of 911 calls in this city originate from neighborhoods in which at least half of the population is Hispanic. However, within those neighborhoods, minority (i.e., mostly Hispanic) and White officers are dispatched to neighborhoods that range from 50 to 100 percent Hispanic.

The correlation between call characteristics and officer race are shown in online Appendix Table A9, which follows the form of Table 2. Panel A shows unconditional correlations, while panels B and C include beat and beat-by-time fixed effects respectively. Estimates in panel A suggest there is more nonrandom sorting of officers in this city across police beats compared to the first city: 6 of the 12 estimates in panel A are significant at the 10 percent level, and 5 are significant at the 1 percent level. However, panels B and C show that once we condition on beat fixed effects, or beat-by-time fixed effects, there is little correlation between officer race and neighborhood and call characteristics. In each panel, only 2 of the 12 coefficients are significant at the 10 percent level, 1 at the 5 percent level, and 1 at the 1 percent level. Moreover, even the largest and most significant coefficients are of small economic significance. For example, the coefficient of 0.000494 in column 5 of panel B suggests that White officers are dispatched to neighborhoods where the unemployment rate is 0.05 percentage points higher. Collectively, results in online Appendix Table A9 indicate that while including beat or beat-by-time fixed effects may not be so necessary in the first city, it appears necessary in the second city. Indeed, given

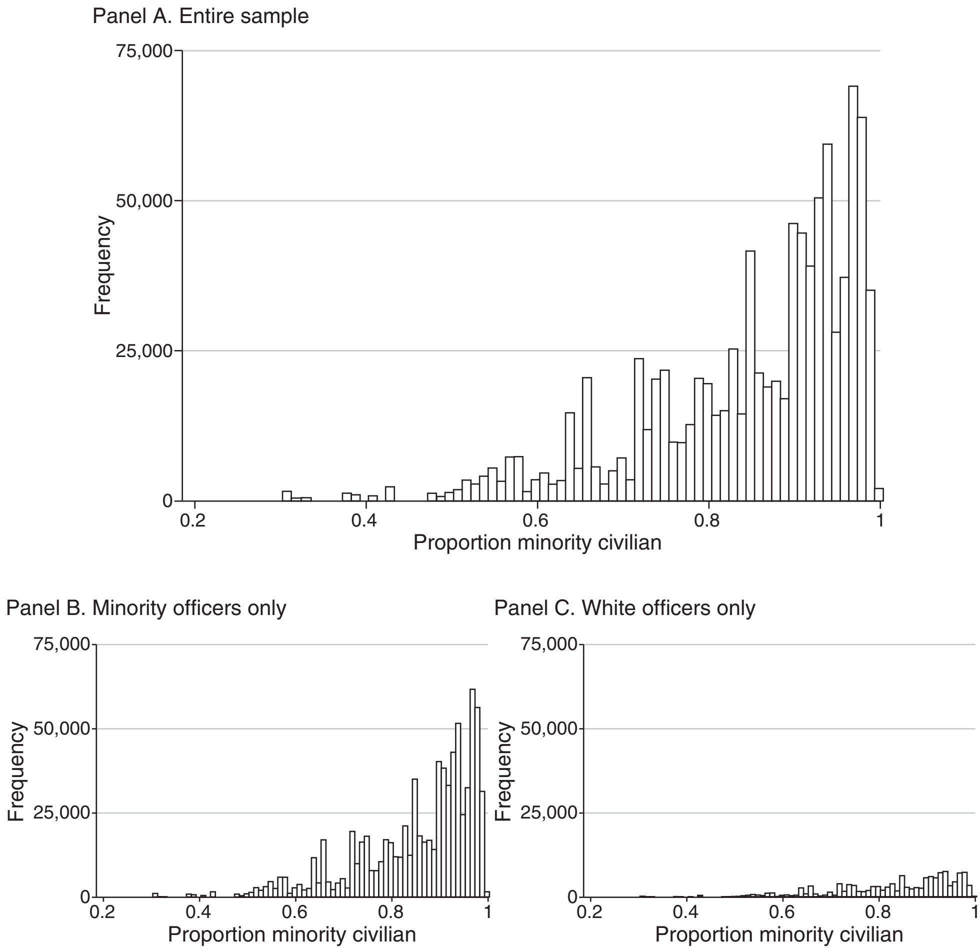

Figure 6. **Second City** Distribution of 911 Calls across Census Block Groups

Notes: These figures report the distribution of proportion minority civilians for calls for service. Each histogram uses 0.01 size bins. Panels B and C report the histograms for calls where only minority or White officers are dispatched, respectively. Ninety-seven percent of minority officers are Hispanic. Ninety-six percent of minority civilians are Hispanic.

the dispatch protocol in the second city, we would expect that controlling for only beat fixed effects would be sufficient in the absence of systematic sorting of officers by race within beat over different time periods within neighborhoods of the same racial composition, and beat-by-time fixed effects should be sufficient under any such nonrandom sorting.

We also graph predicted use of force against the proportion of minorities in the census block group. Specifically, we regress residual use of force (after removing beat-by-time fixed effects) on every call characteristic that we observe, except for the race of the officer dispatched. Results in Figure 7 indicate that White and Hispanic officers were dispatched to calls that had a similar underlying level of danger. This provides evidence that the variation in officer race across calls is as good as random, consistent with the identifying assumption.

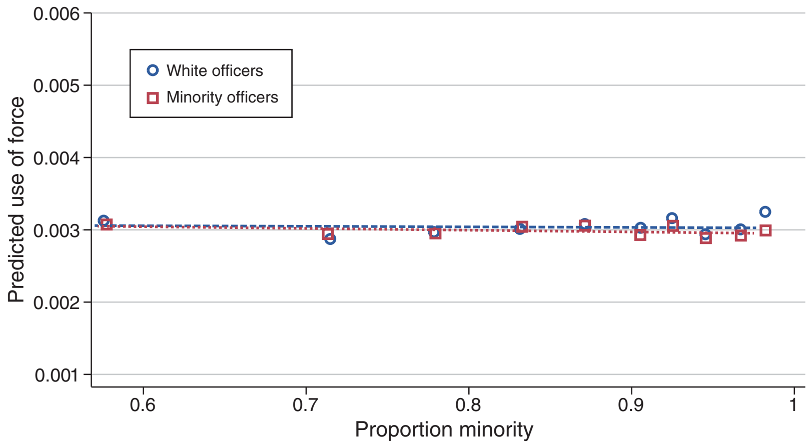

Figure 7. **Second City** Predicted Use of Force for Minority and White Officers

Notes: Here we predict probability of use of force using all observable call characteristics for each call for service. Specifically, we predict (after removing beat-year-month and beat-shift) using proportion minority civilians, call priority, latitude, longitude, per capital income, unemployment, and proportion with less than a high school degree; as well as fixed effects for day of the week, call hour, call description, officer's home beat, and call source. Observations are grouped so that each point includes an equal number of calls. The fitted line is a linear fit across all predicted use of force rates. Ninety-seven percent of minority officers are Hispanic. Ninety-six percent of minority civilians are Hispanic.

Figure 8 shows actual use of force by officer race, though we note it must be interpreted with caution given the nonrandom sorting that is clear from online Appendix Table A9. Two main findings are evident in Figure 8. The first is that White and Hispanic officers do not seem to differ in their overall propensity to use force. This contrasts with Figure 3, which showed that White officers were much more likely to use force than Black officers. The second is that as officers are dispatched to more minority/Hispanic neighborhoods, White officers seem to increase their use of force more than Hispanic officers.

Table 5 shows estimates from specifications that control for beat fixed effects (odd columns) or beat-by-time and call characteristics (even columns).[22] Estimates in columns 1 and 2 indicate White officers use force at roughly the same rate as Hispanic officers. Both coefficients are insignificant, and are economically small in magnitude at 7.6 to 8.2 percent of the mean. However, columns 3 and 4 indicate this overall similarity in use of force disguises the fact that White officers are more likely to use force in Hispanic neighborhoods than White ones. Both estimates are significant at the 5 percent level and indicate White officers are using roughly twice the average amount of force in Hispanic neighborhoods. This previews our main difference-in-difference estimates in columns 5 and 6, which as in Table 3 each

[22] We do not report bootstrapped standard errors here because we cannot do so while weighting each call by the inverse of the number of officers. However, we note that the unweighted results corresponding to Table 5, clustered standard errors are nearly identical to the block bootstrapped standard errors. Specifically, across specifications corresponding to columns 1–6 of Table 5, clustered (bootstrapped) standard errors are 0.000214 (0.000212) and 0.000166 (0.000165) for the indicator on White officer in columns 1 and 2 and are 0.00103 (0.00103), 0.000901(0.000917), 0.00107 (0.00108), and 0.000918 (0.000931) for the interaction term in columns 3–6, respectively.

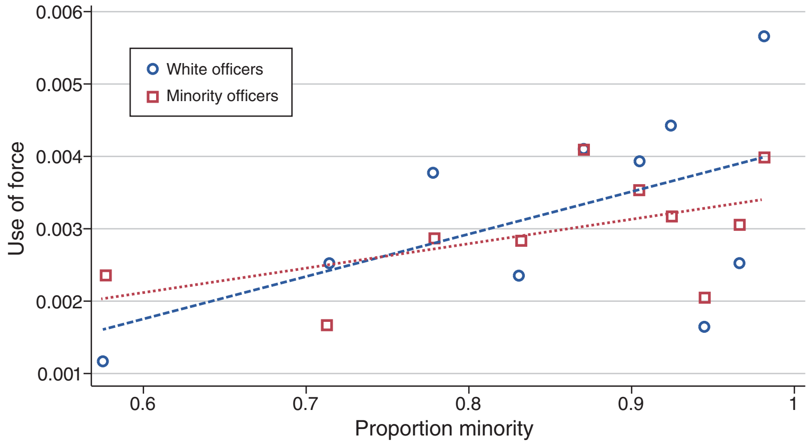

Figure 8. **Second City** Actual Use of Force for Minority and White Officers

Notes: Here we plot the average use of force for ten bins in panel A. The fitted line is a linear fit across all use of force rates. Observations are grouped so that each point includes an equal number of calls. Ninety-seven percent of minority officers are Hispanic. Ninety-six percent of minority civilians are Hispanic.

control for individual officer fixed effects. Estimates indicate that White officers scale up use of force at a rate that is 76 percent (column 6) to 98 percent (column 5) of the mean rate of use of force. Only the latter is significant at the 5 percent level, though both estimates are significant at the 10 percent level.[23]

Online Appendix Table A10 shows the robustness of the estimates in Table 5 to estimation using logit and to the inclusion of interactions. The format of online Appendix Table A10 follows that of Table 4 for the first city. Results from logit estimation of the individual fixed effects specification shown in column 6 suggests an effect of similar magnitude (odds ratio = 1.68), albeit one that is not statistically significant. The estimate in column 7 shows that estimated difference-in-difference effect is similar when we include interactions between call characteristics and officer race and other officer characteristics and civilian race. The estimate is 0.00858 and is significant at the 5 percent level, which is similar to the corresponding estimate in Table 5 of 0.000925.

In summary, our analysis of the second city that is populated primarily by Hispanics and Whites yields two findings. First, White and Hispanic officers use force at similar overall rates. Second, that overall similarity in the use of force disguises the fact that force is disproportionately used in different-race neighborhoods. Results suggest that the rate at which White officers use force increases by more as those officers are dispatched to more Hispanic neighborhoods, compared to Hispanic officers. As a result, we estimate that minority citizens are 75–100 percent as likely to experience use of force when interacting with a White officer, compared to if White officers were to scale up use of force similarly to Hispanic officers.

[23] Two-sided randomization p-values for the interactions in columns 3–6 are 0.011, 0.04, 0.014, and 0.063, respectively.

IV. Discussion and Conclusion

In this paper, we examine whether officer and civilian race matter when it comes to police use of force. We do so by exploiting as-good-as-random variation in the race of officers dispatched to more than two million 911 calls in two different cities. In doing so, we answer two questions: Do White officers use force more often than minority officers in otherwise similar situations? And do White officers scale up their use of force more than minority officers as they go from White to minority neighborhoods?

Results provide strong evidence that race matters. We show that White officers use force 60 percent more than Black officers when responding to similar calls. This increased force is driven primarily by the use of additional force in Black neighborhoods. Difference-in-difference estimates indicate White officers scale up use of force 55 percent more than Black officers as they go from White to Black neighborhoods. Results for use of force with a gun are even more striking. While White and Black officers fire their guns at similar rates in White neighborhoods, White officers are five times more likely than Black officers to use force in Black neighborhoods. Results from a second city shows even larger differences in how White and Hispanic officers scale up force across neighborhoods. Difference-in-difference estimates indicate that officers are 75–100 percent as likely to use force in different-race neighborhoods.

We also perform an analysis of individual officers in the first city. The primary advantage of doing so in this setting is we are able to overcome the problem of selection, which is the major impediment to correctly identifying officers in the tail of any distribution. Our analysis flags 13 percent of White officers but only 5 percent of Black officers as being in the right tail of the use of force distribution. Similarly, we flag 14 percent of White officers and 4 percent of Black officers as being in the right tail of the distribution in Black neighborhoods. We also show that the average racial differences in use of force are driven by a sizable proportion of officers. We would need to drop 32 percent of White officers to eliminate the overall difference in use of force between Black and White officers, and 31 percent of White officers to eliminate the difference-in-difference estimate.

We note that our findings on officer-involved shootings contrasts significantly with previous work by Fryer (2019) and Weisburst (2019), who find Black civilians are no more likely to experience use of force with a gun. There are several explanations for the difference. The first is that the populations and contexts studied are different. In particular, as shown in Figure 4, our findings on force with a gun are driven by police behavior in the predominantly Black neighborhoods of the city, where a disproportionate amount of crime occurs and 911 calls originate. In addition, the city itself has a homicide rate that ranks in the top 20 nationwide among large cities, a set that does not include any of the cities studied by this previous work. This suggests that just as effects are not evident in many of the neighborhoods shown in Figure 4, effects may also not be present in the Black neighborhoods found in less dangerous cities.

Alternatively, the difference in findings could be due to differences in research design. For example, Fryer imposes the benchmarking assumption that encounters that ended in certain types of arrest or nonlethal force were a good approximation of interactions that could have ended in use of force with a gun, but did not. However, if officers are more likely to escalate a situation with a Black civilian versus an otherwise similar White one, or if the officer is more likely to arrest a Black civilian than an otherwise similar White civilian, that benchmarking assumption would lead to understated effects. The same would be true in our research design: if White officers are more likely than Black officers to arrest an otherwise similar civilian, then conditioning on arrest would necessarily attenuate the effect we document. Finally, it is also possible that Fryer (2019)'s selection-on-observables approach could understate the impact of race compared to the difference-in-difference approach used here, perhaps because the context of the encounter is described in the arrest report after force is used.

For these reasons, in addition to providing credible evidence on the extent to which race matters in police use of force in two cities, we believe an important contribution of our study is to provide a clear road map for how to answer that question in other cities without imposing selection-on-observables or benchmarking assumptions. While the 911 call dispatch system in some cities may not be suitable for an analysis like this, others certainly are, and nearly every system could be if policymakers desired. In those cases, the only roadblock to documenting the impact of race on force in this way is the political will to provide access to the necessary data and perform the analysis. Similarly, the approach used here can also be used to track the performance of individual officers without resorting to matching and other selection-on-observables-based techniques to address the problem of selection.

Overall, our findings indicate that race is an important determinant of use of force. While it is difficult to know if these findings extend beyond the two major cities studied here, we note that in both cities White officers scale up force more than minority officers do as they go to minority neighborhoods, and by similar amounts (55 versus 75–100 percent). That is especially striking given the cities differ along many dimensions, not least of which is the racial composition of the cities. The differences in use of force by race matter significantly. Results in both cities suggest that a reallocation of White and minority officers to calls from same-race neighborhoods would reduce overall use of force and officer-involved shootings. Our model estimates that a reshuffling of officers across calls by race in the first city would reduce overall force and officer-involved shootings by 32 and 72 percent in neighborhoods that are over 75 percent Black, respectively, and by 6 percent and 28 percent overall. This reassignment would also reduce the overall racial gap in police use of force per capita by 68 percent, and completely eliminate the racial gap in officer-involved shootings.[24]

Moreover, these findings corroborate the distrust of police among minorities in the United States. The central findings raise two important issues. The first is that we find evidence that the major city we study seems to attract White candidates for the police force who have a much higher propensity to use force than Black

[24] Source: We estimate a model using all covariates, including officer race and the interaction between officer race and proportion Black, to predict use of force given the current assignment of officers. We then predict use of force if we were to assign Black officers to the same number of calls, starting with calls in the most Black neighborhoods. Under this (admittedly extreme) reshuffling, Black officers are only dispatched to calls in neighborhoods that are at least 82 percent Black, and White officers to calls that are less than 82 percent Black. The racial gap in use of force is defined as the number of incidents per 100,000 population in neighborhoods that are over 75 percent Black minus the number per 100,000 population in neighborhoods that are less than 25 percent Black.

candidates. The second is that while use of force rates in Black neighborhoods are usually higher than in White neighborhoods, it is hard to rule out the possibility that more force was merited. The central finding of this paper raises the important question: If calls in minority neighborhoods were that much more dangerous than calls in White neighborhoods, why don't minority officers increase their use of force as much as White officers do? Our findings are inconsistent with the view that race is not a factor in use of force, or that only a handful of officers drive racial effects. Rather, we believe the only reasonable interpretation is that race matters in a systematic way with respect to police use of force.

REFERENCES

Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91 (1): 163–77.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–51.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *Quarterly Journal of Economics* 127 (2): 1017–55.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133 (4): 1885–1932.

Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review* 85 (3): 304–21.

Ba, Bocar A., Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702.

Barsamian Kahn, Kimberly, and Paul G. Davies. 2017. "What Influences Shooter Bias? The Effects of Suspect Race, Neighborhood, and Clothing on Decisions to Shoot." *Journal of Social Issues* 73 (4): 723–43.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.

Bielen, Samantha, Wim Marneffe, and Naci H. Mocan. 2018. "Racial Bias and In-group Bias in Judicial Decisions: Evidence from Virtual Reality Courtrooms." NBER Working Paper 25355.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47 (1): 225–38.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Dahl, Gordon B., and Enrico Moretti. 2008. "The Demand for Sons." *Review of Economic Studies* 75 (4): 1085–1120.

Efron, Bradley. 2004. "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis." *Journal of the American Statistical Association* 99 (465): 96–104.

Feld, Jan, Nicolás Salamanca, and Daniel S. Hamermesh. 2016. "Endophilia or Exophobia: Beyond Discrimination." *Economic Journal* 126 (594): 1503–27.

Flanagan, Francis X. 2018. "Race, Gender, and Juries: Evidence from North Carolina." *Journal of Law and Economics* 61 (2): 189–214.

Fryer, Roland G. 2019. "An Empirical Analysis of Racial Differences in Police Use of Force." *Journal of Political Economy* 127 (3): 1210–61.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." *American Economic Review* 90 (4): 715–41.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111 (5): 1406–41.

Grogger, Jeffrey, and Greg Ridgeway. 2006. "Testing for Racial Profiling in Traffic Stops from behind a Veil of Darkness." *Journal of the American Statistical Association* 101 (475): 878–87.

Haseman, Janie, Karina Zaiets, Mitchell Thorson, Carlie Procell, George Petras, and Shawn J. Sullivan. 2020. "Tracking Protests Across the USA in the Wake of George Floyd's Death." *USA Today*, June 3. https://www.usatoday.com/in-depth/graphics/2020/06/03/map-protests-wake-george-floyds-death/ 5310149002/.

Heckman, James J. 1981. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence."

In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press.

Hoekstra, Mark, and CarlyWill Sloan. 2022. "Replication Data for: Does Race Matter for Police Use of Force? Evidence from 911 Calls." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E143541V1.

Hoekstra, Mark, and Brittany Street. 2021. "The Effect of Own-Gender Jurors on Conviction Outcomes." *Journal of Law and Economics* 64 (3): 513–37.

Horrace, William C., and Shawn M. Rohlin. 2016. "How Dark Is Dark? Bright Lights, Big City, Racial Profiling." *Review of Economics and Statistics* 98 (2): 226–32.

Johnson, David J., Trevor Tress, Nicole Burkel, Carley Taylor, and Joseph Cesario. 2019 Retraction published July 2020. "Officer Characteristics and Racial Disparities in Fatal Officer-Involved Shootings." *Proceedings of the National Academy of Sciences* 116 (32): 15877–82.

Knox, Dean, and Jonathan Mummolo. 2019. "Making Inferences about Racial Disparities in Police Violence." SSRN 3431132.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2019. "The Bias Is Built in: How Administrative Records Mask Racially Biased Policing." SSRN 3336338.

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92 (10–11): 2083–105.

Mekawi, Yara, and Konrad Bresin. 2015. "Is the Evidence from Racial Bias Shooting Task Studies a Smoking Gun? Results From a Meta-Analysis." *Journal of Experimental Social Psychology* 61: 120–30.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences* 109 (41): 16474–79.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics* 111 (3): 915–41.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." *American Economic Review* 101 (4): 1410–35.

Persico, Nicola, and Petra Todd. 2006. "Generalising the Hit Rates Test for Racial Bias in Law Enforcement, with an Application to Vehicle Searches in Wichita." *Economic Journal* 116 (515): F351–67.

Pew Research Center. 2016. *The Racial Confidence Gap in Police Performance*. Washington, DC: Pew Research Center.

Price, Joseph, and Justin Wolfers. 2010. "Racial Discrimination among NBA Referees." *Quarterly Journal of Economics* 125 (4): 1859–87.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122 (6): 1320–54.

Ridgeway, Greg. 2016. "Officer Risk Factors Associated with Police Shootings: A Matched Case–Control Study." *Statistics and Public Policy* 3 (1): 1–6.

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 (486): 661–68.

Sloan, CarlyWill. 2019. "Racial Bias by Prosecutors: Evidence from Random Assignment." Unpublished.

Tuttle, Cody. 2019. "Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums." Unpublished.

United States Census Bureau. 2010. "Table 3 (Race)." Decennial Census. (accessed 2018).

United States Census Bureau. 2016a. "B15003 (Educational Attainment for the Population 25 Years and Older)." American Community Survey 5-Year Data. (accessed 2018).

United States Census Bureau. 2016b. "B19301 (Per Capita Income in the Past 12 Months)." American Community Survey 5-Year Data. (accessed 2018).

United States Census Bureau. 2016c. "B23025 (Employment status for the population 16 and older)." American Community Survey 5-year. (accessed 2018).

Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. "Language from Police Body Camera Footage Shows Racial Disparities in Officer Respect." *Proceedings of the National Academy of Sciences* 114 (25): 6521–26.

Weisburst, Emily K. 2017. "Whose Help Is on the Way? The Importance of Individual Police Officers in Law Enforcement Outcomes." Unpublished.

Weisburst, Emily K. 2019. "Police Use of Force as an Extension of Arrests: Examining Disparities Across Civilian and Officer Race." *AEA Papers and Proceedings* 109: 152–56.

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

This article has been cited by:

1. Peter A. Hanink, Matthew Renner, Christopher J. Bates. 2023. Lethality of Police Shootings and Proximity to Trauma Centers. *Homicide Studies* **27**:1, 120-141. [Crossref]
2. Kristi L. Allgood, Nancy L. Fleischer, Jeffrey Morenoff, Shervin Assari, Belinda L. Needham. 2023. Do Police Encounters Increase the Risk for Cardiovascular Disease? Police Encounters and Framingham 30-Year Cardiovascular Risk Score. *Journal of Racial and Ethnic Health Disparities* **43**. . [Crossref]
3. David Bjerk. 2022. Does greater police funding help catch more murderers?. *Journal of Empirical Legal Studies* **19**:3, 528-559. [Crossref]
4. Tzipporah P. Dang, Bradley D. Mattan, Denise M. Barth, Grace Handley, Jasmin Cloutier, Jennifer T. Kubota. 2022. Perceiving social injustice during arrests of Black and White civilians by White police officers: An fMRI investigation. *NeuroImage* **255**, 119153. [Crossref]
5. Anna Harvey, Taylor Mattia. 2022. Reducing racial disparities in crime victimization: Evidence from employment discrimination litigation. *Journal of Urban Economics* **85**, 103459. [Crossref]
6. Stephanie Holmes Didwania. 2022. Gender Favoritism among Criminal Prosecutors. *The Journal of Law and Economics* **65**:1, 77-104. [Crossref]
7. Deepak Premkumar. 2020. Intensified Scrutiny and Bureaucratic Effort: Evidence from Policing and Crime After High-Profile, Officer-Involved Fatalities. *SSRN Electronic Journal* **103**. . [Crossref]
8. Elizabeth Luh. 2019. Not So Black and White: Uncovering Racial Bias from Systematically Masked Police Reports. *SSRN Electronic Journal* **25**. . [Crossref]