

An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On non-lethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From “Bloody Sunday” on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019

[*Journal of Political Economy*, 2019, vol. 127, no. 3]

© 2019 by The University of Chicago. All rights reserved. 0022-3808/2019/12703-0006\$10.00

between African Americans and police has an unlovely history. The images of law enforcement clad in Ku Klux Klan regalia or peaceful protesters being attacked by canines, high-pressure water hoses, and tear gas are an indelible part of American history. For much of the twentieth century, law enforcement chose to brazenly enforce the status quo of overt discrimination rather than protect and serve all citizens.

The raw memories of these injustices have been resurrected by several high-profile incidents of questionable uses of force. Michael Brown, unarmed, was shot 12 times by a police officer in Ferguson, Missouri, after Brown fit the description of a robbery suspect of a nearby store. Eric Garner, unarmed, was approached because officers believed he was selling single cigarettes from packs without tax stamps, and in the process of arresting him, an officer choked him and he died. Walter Scott, unarmed, was stopped because of a nonfunctioning third brake light and was shot eight times in the back while attempting to flee. Samuel Du Bose, unarmed, was stopped for failure to display a front license plate and while trying to drive away was fatally shot once in the head. Rekia Boyd, unarmed, was killed by a Chicago police officer who fired five times into a group of people from inside his police car. Zachary Hammond, unarmed, was driving away from a drug deal sting operation when he was shot to death by a Seneca, South Carolina, police officer. He was white. And so are 44 percent of police shooting subjects.¹

These incidents, some captured on video and viewed widely, have generated protests in Ferguson, New York City, Washington, Chicago, Oakland, California, and several other cities and a national movement (Black Lives Matter) and a much-needed national discourse about race, law enforcement, and policy. Police precincts from Houston to Camden, New Jersey, to Tacoma, Washington, are beginning to issue body-worn cameras, engaging in community policing, and enrolling officers in training in an effort to purge racial bias from their instinctual decision making. However, for all the eerie similarities between the current spate of police interactions with African Americans and the historical injustices that remain unhealed, the current debate is virtually data-free. Understanding the extent to which there are racial differences in police use of force and (if any) whether those differences might be due to discrimination by police

School of Economics, University College London, and the NBER Summer Institute for helpful comments and suggestions. Brad Allan, Elijah De La Campa, Tanaya Devi, William Murdock III, and Hannah Ruebeck provided truly phenomenal project management and research assistance. Lukas Althoff, Dhruva Bhat, Samarth Gupta, Julia Lu, Mehak Malik, Beatrice Masters, Ezinne Nwankwo, Sofya Shchukina, and Eric Yang provided excellent research assistance. Financial support from the EdLabs Advisory Group and an anonymous donor is gratefully acknowledged. The usual caveat applies. Data set availability is described in supplementary material provided online.

¹ Author's calculations based on ProPublica research that analyzes FBI data between 1980 and 2012.

or explained by other factors at the time of the incident is a question of tremendous social importance and is the subject of this paper.

A primary obstacle to the study of police use of force has been the lack of readily available data. Data on lower-level uses of force, which happen more frequently than officer-involved shootings, are virtually nonexistent. The reason in part is that most police precincts do not explicitly collect data on use of force and in part that even when the data are hidden in plain view within police narrative accounts of interactions with civilians, they are exceedingly difficult to extract. Moreover, the task of compiling rich data on officer-involved shootings is burdensome. Until recently, data on officer-involved shootings were extremely rare and contained little information on the details surrounding an incident. A simple count of the number of police shootings that occur does little to explore whether racial differences in the frequency of officer-involved shootings are due to police malfeasance or differences in suspect behavior.²

In this paper, we estimate the extent of racial differences in police use of force using four separate data sets—two constructed for the purposes of this study.³ Unless otherwise noted, all results are conditional on an interaction. Understanding potential selection into police data sets due to bias in whom police interact with is a difficult endeavor. Section III attempts to help get a sense of potential bias in police interactions. Put simply, if one assumes that police simply stop whomever they want for no particular reason, there seem to be large racial differences. If one assumes they are trying to prevent violent crimes, then evidence for bias is small.

Of the four data sets, the first comes from NYC's Stop, Question, and Frisk program (hereafter Stop and Frisk). Stop and Frisk is a practice of the New York City Police Department in which police stop and question pedestrians and then can frisk them for weapons or contraband. The data set contains roughly 5 million observations. And, important for the purposes of this paper, it has detailed information on a wide range of uses of force—from putting hands on civilians to striking them with a baton. The second data set is the Police-Public Contact Survey, a triennial survey of a nationally representative sample of civilians, which contains—from the civilian point of view—a description of interactions with police, which includes uses of force. Both these data sets are public-use and readily available.⁴

² Newspapers such as the *Washington Post* (Kindy et al. 2015) estimate that there were almost 1,000 officer-involved shootings in 2015. Websites such as Fatal Encounters (<http://www.fatalencounters.org>) estimate that the number of annual shootings is approximately 704 between 2000 and 2015.

³ Throughout the text, I depart from custom by using the terms “we,” “our,” and so on. Although this is a sole-authored work, it took a large team of talented individuals to collect the data necessary for this project. Using “I” seems disingenuous.

⁴ The NYC Stop and Frisk data have been used in Gelman, Fagan, and Kiss (2012) and Coviello and Persico (2015) to understand whether there is evidence of racial discrimina-

The other two data sets were assembled for the purposes of this research. We use event summaries from all incidents in which an officer discharges his weapon at civilians—including both hits and misses—from three large cities in Texas (Austin, Dallas, and Houston), Denver, Seattle, nine large Florida counties, Jacksonville, and Los Angeles County to construct a data set in which one can investigate racial differences in officer-involved shootings. Because all individuals in these data have been involved in a police shooting, analysis of these data alone can estimate racial differences only on the intensive margin (e.g., whether the officers discharged their weapons before or after the suspect attacked).

To supplement, our fourth data set contains a random sample of police-civilian interactions from the Houston Police Department from arrest codes in which lethal force is more likely to be justified: attempted capital murder of a public safety officer, aggravated assault on a public safety officer, resisting arrest, evading arrest, and interfering in an arrest. Similar to the event studies above, these data come from arrest narratives that range in length from two to 100 pages. A team of researchers was responsible for reading arrest reports and collecting almost 300 variables on each incident. Combining this with the officer-involved shooting data from Houston allows us to estimate both the extensive (e.g., whether or not a police officer decides to shoot) and intensive margins. Further, the Houston arrest data contain almost 4,500 observations in which officers discharged charged electronic devices (e.g., tasers). This is the second most extreme use of force and, in some cases, is a substitute for lethal force.

The results obtained using these data are informative and, in some cases, startling. Using data on police interactions from NYC's Stop and Frisk program, we demonstrate that on nonlethal uses of force—putting hands on civilians (which includes slapping or grabbing) or pushing individuals into a wall or onto the ground—there are large racial differences. In the raw data, blacks and Hispanics are more than 50 percent more likely to have an interaction with police that involves any use of force. Accounting for 125 variables that represent baseline characteristics, encounter characteristics, civilian behavior, precinct, and year fixed effects, the odds ratio for blacks (respectively, Hispanics) is 1.178 (respectively, 1.122).

Interestingly, as the intensity of force increases (e.g., handcuffing civilians without arrest, drawing or pointing a weapon, or using pepper spray or a baton), the probability that any civilian is subjected to such treatment is small, but the racial difference remains surprisingly constant. For instance, 0.26 percent of interactions between police and civilians involve

tion in proactive policing, and in Ridgeway and MacDonald (2009) to develop a statistical method to identify problem officers. The Police-Public Contact Survey has been used, mainly in criminology, to study questions such as whether police treatment of citizens affects the broader public opinion of the police (Miller et al. 2004).

an officer drawing a weapon; 0.02 percent involve using a baton. These are rare events. Yet the results indicate that they are significantly more rare for whites than for blacks. With all controls, blacks are 21 percent more likely than whites to be involved in an interaction with police in which at least a weapon is drawn, and the difference is statistically significant. Across all nonlethal uses of force, the odds ratio of the black coefficient ranges from 1.175 (0.036) to 1.275 (0.131).

Data from the Police-Public Contact Survey (PPCS) are qualitatively similar to the results from Stop and Frisk data, in terms of both whether or not any force is used and the intensity of force, though the estimated racial differences are significantly larger. Blacks and Hispanics are approximately 1.3 percentage points more likely than whites to report any use of force in a police interaction, including controls for civilian demographics, civilian behavior, contact characteristics, officer characteristics, and year. The white mean is 0.7 percent. Thus, the odds ratio is 2.769 for blacks and 1.818 for Hispanics.

There are several potential explanations for the quantitative differences between our estimates using Stop and Frisk data and those using PPCS data. First, we estimate odds ratios, and the baseline probability of force in each of the data sets is substantially different. Second, the PPCS is a nationally representative sample of a broad set of police-civilian interactions. Stop and Frisk data are from a particular form of policing in a dense urban area. Third, the PPCS is gleaned from the civilian perspective. Finally, granular controls for location are particularly important in the Stop and Frisk data and unavailable in PPCS. In the end, the “answer” is likely somewhere in the middle, and, importantly, both bounds are statistically and economically important.

In stark contrast to nonlethal uses of force, we find that, conditional on a police interaction, there are no racial differences in officer-involved shootings on either the extensive or intensive margin. Using data from Houston—where we have both officer-involved shootings and a randomly chosen set of potential interactions with police in which lethal force may have been justified—we find, after controlling for suspect demographics, officer demographics, encounter characteristics, suspect weapon, and year fixed effects, that blacks are 27.4 percent less likely to be shot at by police relative to nonblack, non-Hispanics. This coefficient is measured with considerable error and is not statistically significant. This result is remarkably robust across alternative empirical specifications and subsets of the data. Partitioning the data in myriad ways, we find no evidence of racial discrimination in officer-involved shootings. Investigating the intensive margin—the timing of shootings or how many bullets were discharged in the endeavor—there are no detectable racial differences.

Our results have several important caveats. First, all but one data set were provided by a select group of police departments. It is possible that these

departments supplied the data only because they either are enlightened or were not concerned about what the analysis would reveal. In essence, this is equivalent to analyzing labor market discrimination on a set of firms willing to supply a researcher with their human resources data! There may be important selection in who was willing to share their data. The PPCS partially sidesteps this issue by including a nationally representative sample of civilians, but it does not contain data on officer-involved shootings. Moreover, Fryer (2018) demonstrates that, if anything, our data seem to contain more “antiblack bias” when analyzed using the descriptive statistics deployed in Kindy et al. (2015), Ross (2015), or Swaine et al. (2015).

Relatedly, even police departments willing to supply data may contain police officers who present contextual factors at the time of an incident in a biased manner—making it difficult to interpret regression coefficients in the standard way.⁵ It is exceedingly difficult to know how prevalent this type of misreporting bias is (Schneider 1977). Accounting for contextual variables recorded by police officers who may have an incentive to distort the truth is problematic. Yet, whether or not we include controls does not alter the basic qualitative conclusions. And, to the extent that there are racial differences in underreporting of nonlethal use of force (and police are more likely to not report force used on blacks), our estimates may be a lower bound. Not reporting officer-involved shootings seems unlikely.

Third, given the inability to randomly assign race, one can never be confident in the direct regression approach when interpreting racial disparities. We partially address this in two ways. First, we build a model of police-civilian interactions that allows for both statistical and taste-based discrimination and use the predictions of the model to help interpret the data. For instance, if police officers are pure statistical discriminators, then as civilians’ signal to police regarding their likelihood of compliance becomes increasingly deterministic, racial differences should disappear. To test this, we investigate racial differences in use of force on a set of police-civilian interactions in which the police report that the civilian was compliant on every measured dimension, he was not arrested, and neither weapons nor contraband was found. In contrast to the model’s predictions, racial differences on this set of interactions are large and statistically significant. Additionally, we demonstrate that the *marginal* returns to compliant behavior are the same for blacks and whites, but the *average*

⁵ In the Samuel DuBose case at the University of Cincinnati, the officer reported, “Mr. DuBose pulled away and his arm was caught in the car and he got dragged,” yet body camera footage showed no such series of events. In the Laquan McDonald case in Chicago, the police reported that McDonald lunged at the officer with a knife while dash-cam footage showed the teenager walking away from the police with a small knife when he was fatally shot 16 times by the officer.

return to compliance is lower for blacks—suggestive of taste-based, rather than statistical, discrimination.

For officer-involved shootings, we employ a simple Beckerian outcomes test (Becker 1993) for discrimination inspired by Knowles, Persico, and Todd (2001) and Anwar and Fang (2006). We investigate the fraction of white and black suspects, separately, who are armed conditional on being involved in an officer-involved shooting. If the ordinal threshold of shooting at a black suspect versus a white suspect is different across officer races, then one could reject the null hypothesis of no discrimination. Our results, if anything, are the opposite. We cannot reject the null of no discrimination in officer-involved shootings.

Taken together, we argue that the results are most consistent with, but in no way proof of, taste-based discrimination among police officers who face convex costs of excessive use of force. Yet the data do more to provide a more compelling case that there is no discrimination in officer-involved shootings than they do to illuminate the reasons behind racial differences in nonlethal uses of force.

The rest of the paper is organized as follows. The next section describes and summarizes the four data sets used in the analysis. Section III describes potential selection into police data sets. Section IV presents estimates of racial differences on nonlethal uses of force. Section V describes a similar analysis for the use of lethal force. Section VI attempts to reconcile the new facts with a simple model of police-civilian interaction that incorporates both statistical and taste-based channels of discrimination. The final section presents conclusions. There are three online appendixes. Appendix A describes the data used in our analysis and how we coded variables. Appendix B describes the process of creating data sets from event summaries. Appendix C provides additional theoretical results.

II. The Data

We use four sources of data—none ideal—which together paint an empirical portrait of racial differences in police use of force conditional on an interaction. The first two data sources—NYC’s Stop, Question, and Frisk program and the Police-Public Contact Survey (PPCS)—provide information on nonlethal force from both the police and civilian perspectives, respectively. The other two data sets—event summaries of officer-involved shootings in 16 locations across the United States and data on interactions between civilians and police in Houston, in which the use of lethal force may have been justified by law—allow us to investigate racial differences in officer-involved shootings on both the extensive and intensive margins.

Below, I briefly discuss each data set in turn. Appendix A provides further detail.

A. New York City's Stop, Question, and Frisk Program

New York City's Stop, Question, and Frisk data consist of 5 million individual police stops in NYC between 2003 and 2013. The database contains detailed information on the characteristics of each stop (precinct, cross streets, time of day, inside/outside, high-/low-crime area), civilian demographics (race, age, gender, height, weight, build, type of identification provided), whether or not the officers were in uniform, encounter characteristics (reason for stop, reason for frisk [if any], reason for search [if any], suspected crime[s]), and postencounter characteristics (whether or not a weapon was eventually found or whether an individual was summonsed or arrested or whether a crime was committed).

Perhaps the most novel component of the data is that officers are required to document which one of the following seven uses of force was used, if any: (1) put hands on a civilian, (2) force to a wall, (3) handcuff, (4) draw weapon, (5) push to the ground, (6) point a weapon, (7) pepper spray or strike with a baton.⁶ Officers are instructed to include as many uses of force as applicable. For instance, if a stop results in an officer putting his hands on a civilian and, later within the same interaction, pointing his weapon, that observation would have both "hands" and "point a weapon" as uses of force. Unfortunately, officers are not required to document the sequence in which they used force.

These data have important advantages. First, the Stop and Frisk program encompasses a diverse sample of police-civilian interactions.⁷ Between the years 2003 and 2013, the same period as the Stop and Frisk data, there were approximately 3,457,161 arrests in NYC—26.3 percent fewer observations than Stop and Frisk excluding stops that resulted in arrests.⁸ Unfortunately, even this robust data set is incomplete: nowhere is the universe of all police interactions with civilians—or even all police stops—recorded.

Second, lower-level uses of force—such as the use of hands—are both recorded in these data and more frequently used by law enforcement than more intense uses of force. For instance, if one were to use arrest data to glean use of force, many lower-level uses of force would simply be considered standard operating procedure. Putting hands on suspects,

⁶ Police officers can also include "other" force as a type of force used against civilians. We exclude other forces from our analysis. Appendix table 4 calculates racial differences in the use of other force and shows that including these forces does not alter our results.

⁷ Technically, NYC police are required to record a stop only if some force was used or a civilian was frisked or searched, was arrested, or refused to provide identification. Nonetheless, roughly 41 percent of all stops in the database appear to be reported despite not resulting in any of the outcomes that legally trigger the requirement to record the stop.

⁸ This number was calculated from the Division of Criminal Justice Services' record of adult arrests by counties in New York City between 2003 and 2013.

pushing them up against a wall, and putting handcuffs on them are so unnoteworthy in the larger context of an arrest that they are not recorded in typical arrest descriptions. Yet, because proactive policing is a larger and less confrontational portion of police work, these actions warrant data entry.

The key limitation of the data is that they capture only the police side of the story. There have been several high-profile cases of police storytelling that is not congruent with video evidence of the interaction. Another important limitation for inference is that the data do not provide a way to identify officers or individuals. Ideally, one would simply cluster standard errors at the officer level to account for the fact that many data points—if driven by a few aggressive officers—are correlated and classic inference treats them as independent. Our typical regressions cluster standard errors at the precinct level. Appendix table 10 explores the robustness of our results for more disaggregated clusters: precinct \times time of day, block level, and even block \times time of day. Our conclusions are unaffected by any of these alternative ways to cluster standard errors.

Summary statistics for the Stop and Frisk data are displayed in appendix table 2A. There are six panels. Panel A contains baseline characteristics. Fifty-eight percent of all stops recorded were of black civilians. If police were stopping individuals at random, this number would be closer to 25.5 percent (the fraction of black civilians in New York City according to US Census Bureau 2010 records). Hispanics make up 25 percent of the stops. The data comprise predominantly young males; the median age is 24 years old. The median age in NYC is roughly 11 years older.

Panel B describes encounter characteristics for the full sample and then separately by race. Most stops occur outside after the sun has set in high-crime areas. There is a surprisingly small number of stops—about 3 percent—in which the police report finding any weapon or contraband. Panel C displays variables that describe civilian behavior. Approximately 50 percent of stops were initiated because a civilian fit the relevant description of a person of interest, the person was assumed to be a lookout for a crime, or the officers were casing a victim or location.

Panel D contains a series of alternative outcomes such as whether a civilian was frisked, summonsed, or arrested. Panel E provides descriptive statistics for the seven forms of force available in the data. Panel F provides the frequency of missing variables.

B. The Police-Public Contact Survey

The PPCS—a nationally representative sample—has been collected by the Bureau of Justice Statistics (BJS) every 3 years since 1996. The most recent wave publicly available is 2011. Across all years, there are approximately 426,000 observations.

The main advantage of the PPCS data is that, unlike any of our other data sets, they provide the civilian's interpretation of interactions with police. The distinction between PPCS data and almost any other data collected by the police is similar to the well-known differences between certain data in the Uniform Crime Reports (UCR) and the National Crime Victimization Survey (NCVS).⁹ One explanation for these differences given in the literature is that individuals are embarrassed or afraid to report certain crimes to police or believe that reporting such crimes has unclear benefits and potential costs. Reporting police use of force—in particular for young minority males—may be similar.

Another key advantage is that the data approximate the universe of potential interactions with police—rather than being limited to arrests or police stops.¹⁰ If a police officer is investigating a crime in a neighborhood and he discusses it with a civilian, this type of interaction would be recorded in the PPCS. Or if a police officer used force on a civilian and did not report the interaction, this would not be recorded in police data but would be included in the PPCS.

The PPCS also has important limitations. First, data on individuals' locations are not available to researchers. There are no geographic indicators. Second, the data on contextual factors surrounding the interaction with police or the officer's characteristics are limited. Third, the survey omits individuals who are currently in jail. Fourth, the PPCS includes only the civilian account of the interaction, which could be biased in its own way. In this vein, according to individuals in the PPCS data, only 3.28 percent of them have resisted arrests, and only 11.07 percent of civilians argued when they were searched despite not being guilty of carrying alcohol, drugs, or weapons.

Appendix table 2B presents summary statistics for the PPCS sample with at least one interaction with police. There are six panels. Panel A contains civilian demographics. Blacks constitute roughly 10 percent of the sample; women are 50 percent. The average age is approximately 13 years older than in the Stop and Frisk data. Over 72 percent of the sample reports being employed in the previous week; the average income category

⁹ According to the US Department of Justice, UCR and NCVS measure an overlapping but nonidentical set of crimes. The UCR program's primary objective is to provide a reliable set of criminal justice statistics by compiling data from monthly law enforcement reports or individual crime incident reports transmitted directly to the FBI or to centralized state agencies that then report to the FBI. The BJS, on the other hand, established the NCVS to provide previously unavailable information about crime (including crime not reported to police), victims, and offenders. Therefore, there are discrepancies in victimization rates from the two reports: the UCR reports 89,000 forcible rapes in 2010 while the NCVS reports 203,830 rapes and sexual assaults in 2010.

¹⁰ Contacts exclude encounters with private security guards, police officers seen on a social basis, police officers related to the survey respondents, or any contacts that occurred outside the United States.

in the sample is 2.09. Income is recorded as a categorical variable that is 1 for income levels below \$20,000, 2 for income levels between \$20,000 and \$49,999, and 3 for income levels greater than \$50,000.

Panel B describes self-reported civilian behavior. According to all PPCS survey respondents, only 1.93 percent of civilians disobey police orders, try to get away, resist, argue, or threaten officers when they have some interaction with the police.

Panel C of appendix table 2B includes summary data on the types of contact and officer characteristics. Almost half of the interactions between the public and police are traffic stops, 0.35 percent are from street interactions—including the types of street interaction that may not appear in our Stop and Frisk data—and 44.73 percent are “other,” which include being involved in a traffic accident, reporting a crime, being provided a service by the police, participating in block watch or other anti-crime programs, or being suspected by the police of something or as part of a police investigation. Panel D contains alternative outcomes, and panel E describes the five uses of force available in the data. Panel F provides the frequency of missing variables.

C. *Officer-Involved Shootings*

There are no systematic data sets that include officer-involved shootings (OIS) along with demographics, encounter characteristics, and suspect and police behavior.¹¹ For the purposes of this project, we compile a data set on OIS from 16 locations across America.

To begin, 20 police departments across the country were contacted by the author: Boston, Camden (New Jersey), New York City, Philadelphia, Austin, Dallas, Houston, Los Angeles, 10 Florida counties, Denver, and Seattle.¹² Importantly for thinking about the representativeness of the data, many of these cities were a part of the Obama administration’s Police Data Initiative.¹³ We received data from all but two of these police departments: NYC and Philadelphia.¹⁴ This is likely not a representative set

¹¹ Data constructed by the *Washington Post* have civilian demographic identifiers, weapons carried by civilians, signs of mental illness, and an indicator for threat level but no other contextual information.

¹² Another approach is to request the data from every police department vis-à-vis a Freedom of Information Act request. We attempted this method, but police departments are not obliged to include detailed event summaries. In our experience, the only way to obtain detailed data is to have contacts within the police department.

¹³ The White House launched the Police Data Initiative as a response to the recommendations made by the Task Force on 21st Century Policing. The initiative was created to work with police departments to leverage data on police-citizen interactions (e.g., OIS, use of force, body camera videos, and police stops) to increase transparency and accountability.

¹⁴ Camden and Boston each had one OIS during the relevant time frame, so we did not use their data for this analysis. Camden provided remarkable data on police-civilian interactions, which may be used in future work.

of cities. Appendix table 17 investigates differences between the cities that provided us data and the national average on a variety of dimensions such as population demographics and crime rates. Fryer (2018) analyzes our set of cities relative to more representative, but less detailed, data collected by Kindy et al. (2015), Swaine et al. (2015), and Arthur et al. (2017).

In most cases, OIS data begin as event summaries from all incidents in which a police officer discharged his firearm at civilians (including both hits and misses). These summaries, in many cases, are more than 50-page descriptions of the factors surrounding an officer-involved shooting. Below is an extract from a “typical” summary:

As I pointed my rifle at the vehicle my primary focus was on the male passenger based on the information provided by the dispatcher as the person who had been armed inside the store. As the vehicle was driving past me I observed the male passenger in the truck turn around in the seat, and begin pointing a handgun at me through an open rear sliding glass window. When I observed this I was still yelling at the female to stop the truck! The male suspect appeared to be yelling at me, but I could not hear him. At that point the truck was traveling southbound toward the traffic light on Atlantic Boulevard, and was approximately 30–40 feet away from me. The car had already passed me so the driver was no longer in my line of fire. I could also see my back drop consisted of a wooded area of tall pine trees. It appeared to me at that time that his handgun was moving in a similar fashion of being fired and going through a recoil process, but I could not hear gunshots. Fearing for my life, the lives of the citizens in the area and my fellow officers I began to fire my rifle at the suspect.

To create a data set out of these narratives, a team of research assistants read each summary and extracted data on 65 predetermined variables in six categories: (A) suspect characteristics, (B) suspect weapon(s), (C) officer characteristics, (D) officer response reason, (E) other encounter characteristics, and (F) location characteristics.¹⁵ Suspect characteristics include data on suspect race, age, and gender. Suspect weapon variables consist of dummy variables for whether the suspect used a firearm, sharp object, vehicle, or other objects as a weapon or did not have a weapon at all. Officer characteristics include variables that determine the majority

¹⁵ Appendix B provides a detailed, step-by-step, account of how the OIS data set was created and was explicitly designed to allow researchers to replicate our analysis from the original source materials.

race of the officer unit, whether there were any female officers in the unit, average tenure of the shooting officer, and dummy variables for whether the officer was on duty and was accompanied by two or more officers on the scene. Officer response reason variables determine the reason behind the officer being present at the scene. They include dummy variables on whether the officer was present as a response to a robbery, a violent disturbance, traffic-related stop; was responding to a warrant, any suspicious activity, a narcotics transaction, or a suicide; or was responding because he was personally attacked or for other reasons. Other encounter characteristics gather information on whether the shooting happened during the day or night and a variable that is coded 1 if the suspect attacked the officer or drew a weapon or attempted to draw a weapon on the officer. The variable is coded 0 if the suspect only appeared to have a weapon or did not attack the officer at all. Finally, location characteristics include dummies to represent the jurisdiction that we collected data from. Appendix B contains more details on how the variables were coded.

As a crucial check on data quality, once we coded all OIS data from the event summaries, we wrote appendix B. We then hired eight new research assistants who did not have any involvement in creating the first data set. We provided them the event summaries, appendix B, and extremely minimal instructions—the type of simple clarification that would be provided to colleagues attempting to replicate our work from the source material—and they created a second, independent, data set. All results remain qualitatively unchanged with the alternatively coded data set (see app. tables 15 and 16).¹⁶

The most obvious advantage of the OIS data is the breadth and specificity of information contained in the event summaries. Descriptions of OIS are typically long and quite detailed relative to other police data. A second advantage is that officer-involved shootings are nonsubjective. Unlike lower-level uses of force, whether or not an officer discharges a weapon is not open to interpretation. Officers are also required to document any time they discharge their weapon. Finally, OIS are subject to internal and oftentimes external review.

The OIS data have several notable limitations. Taken alone, officer-involved shootings are the most extreme and least used form of police force and thus, in isolation, may be misleading. Second, the penalties for wrongfully discharging a lethal weapon in any given situation can be life altering; thus, the incentive to misrepresent contextual factors on police reports may be large.¹⁷ Third, we do not typically have the sus-

¹⁶ Thanks to Derek Neal for suggesting this exercise.

¹⁷ From interviews with dozens of current police officers, we gleaned that in most all police shootings—even when fully justified and observed by many—the officer is taken off active duty, pending an investigation.

pect's side of the story, and often there are no witnesses. Fourth, it is impossible to capture all variables of importance at the time of a shooting. Thus, what appears to be discrimination to some may look like mismeasured contextual factors to others.

A final disadvantage, potentially most important for inference, is that all observations in the OIS data are shootings. In statistical parlance, they do not contain the "zeros" (e.g., set of police interactions in which lethal force was justified but not used). To the extent that racial bias is prevalent on the extensive margin—whether or not someone is ever in an officer-involved shooting—these data would not capture it.

We address this concern both directly and indirectly in two ways. First, given the data we have, we investigate the intensive margin by defining our outcome variable as whether or not the officer shoots the suspect before being attacked. Second, we collected unprecedented data from the Houston Police Department on all arrest categories in which officers could have justifiably used lethal force as a way to obtain the zeros. These data are described in the next subsection.

Appendix table 2C displays summary statistics for OIS data, divided into four locations and six categories of data. Column 1 contains observations from the full sample—1,399 shootings between 2000 and 2015.¹⁸ Forty-six percent of officer-involved shootings in our data are blacks, 29 percent are Hispanics, and 25 percent are other, with the majority in that category being whites. Given the spate of video evidence on police shootings—all of which are of blacks—it is a bit surprising that they are less than half of the observations in the data.

Columns 2 and 4 display data from 508 officer-involved shootings with firearms and over 4,000 instances of an officer-involved shooting with a taser, in Houston. Most police officers in the Houston Police Department carry Glock 22, Glock 23, or the Smith & Wesson M&P40 .40 caliber semi-automatic handguns on their dominant side, but many carry an X26 taser on their nondominant side. We exploit this choice problem to understand how real-time police decisions may be correlated with suspect race.

Columns 5–7 contain OIS data from Austin and Dallas, Texas, 10 Florida counties (Brevard, Escambia, Jacksonville, Hardee, Lee, Orange, Palm Beach, Pasco, Pinellas, and Walton), Denver, Seattle, and Los Angeles County. Panel F demonstrates that Houston accounts for 36 percent of all officer-involved shootings. Austin and Dallas, combined, provide almost 20 percent of the data, while Florida provides almost 30 percent of the data. Panel G provides the frequency of missing variables.

¹⁸ We asked for data on all OIS between 2000 and 2015, and police departments replied back with years they had data on. Walton County, Florida, had only 1 year of data, whereas Houston and Orange County provided 16 years of data (2000–2015).

D. Houston Police Department Arrests Data

The most comprehensive set of OIS data is from the Houston Police Department (HPD). For this reason, we contacted the HPD to help construct a set of police-civilian interactions in which lethal force may have been justified. According to chapter 9 of the Texas Penal Code, police officers' use of deadly force is justified "when and to the degree the actor reasonably believes the force is immediately necessary." Below, we describe the task of implementing this obtuse definition in data in an effort to develop a set of police-civilian interactions in which the use of lethal force may have been justified by law.

There are approximately 1 million arrests per year in Houston, 16 million total over the years for which we have OIS data. If the data were more systematically collected, the tasks of creating potential risk sets would be straightforward. Data in the HPD are the opposite: most of them are narrative reports in the form of unstructured blocks of text that one can link to alternative HPD data with unique case identifiers.¹⁹

We randomly sampled 10 percent of case IDs by year from five arrest categories that are more likely to contain incidences in which lethal force was justified: attempted capital murder of a public safety officer, aggravated assault on a public safety officer, resisting arrest, evading arrest, and interfering in an arrest.²⁰ This process narrowed the set of relevant arrests to 16,000 total between 2000 and 2015. Then we randomly sampled 10 percent of these arrest records by year and manually coded 290 variables per arrest record. It took between 30 and 45 minutes per record to manually keypunch and include variables related to specific locations for calls, incidents, arrests, suspect behavior, suspect mental health, suspect injuries, officer use of force, and officer injuries resulting from the encounter.

These data were merged with data on officer demographics and suspects' previous arrest history to produce a comprehensive incident-level data set on interactions between police and civilians in which lethal force may have been justified.

We also collected 4,250 incident reports for all cases in which officers discharged their taser. These data form another potential risk set. It it

¹⁹ In conversations with engineers and data scientists at Google, Microsoft Research, and several others in artificial intelligence and machine learning, we were instructed that current natural language processing algorithms are not developed for the level of complexity in our police data. Moreover, one would need a "test sample" (manually coded data to assess the algorithm's performance) of several hundred thousand to design an algorithm. This is outside the scope of the current project.

²⁰ Our original request to the HPD was for a data set similar to OIS for all arrests between 2000 and 2015. The response was "we estimate that it will take 375 years to fulfill that request."

important to note that technology allows for the HPD to centrally monitor the frequency and location of taser discharges.

Appendix table 2C, column 3, provides descriptive statistics for the Houston arrest data. Compared to the OIS data set, civilians sampled in the arrest data set carry far fewer weapons: 95 percent do not carry weapons compared to 21 percent in the OIS data set. The other variable that is significantly different between the two data sets is the fraction of suspects who attacked or drew a weapon: 56 percent in the HPD arrest data set compared to 80 percent in the OIS data set.

III. A Note on Potential Selection into Police Data Sets

The forthcoming analysis takes the four data sets described above as given and estimates racial differences in nonlethal and lethal uses of force. But, to the extent that there are racial differences in the probability of an interaction with police, these data may omit a very important margin. Put differently, one may discover no differences in police use of force, conditional on an interaction, but large racial differences in the probability of the types of interactions in which force may be used. By concentrating only on how and whether force was used in an interaction and ignoring whether or not an interaction took place, one can misrepresent the total experience with police.

Understanding racial differences in the probability of police interaction is fraught with difficulty. One has to account for differential exposure to police, race-specific crime participation rates, and perhaps, most importantly, pre-interaction behavior that civilians exhibit. Ideally, one might set up a field experiment—similar to those used to measure labor market discrimination—that randomly assigns similar individuals (across all physical dimensions except race) to the vicinity of the same patrolling officers in a neighborhood and instruct them to behave identically. Conditional on random assignment, identical behavior, and race-specific crime rates, any differences in the probability of interaction could be interpreted as racial bias in police stopping behavior.

Without ideal data, researchers often compare the racial distribution of stopped civilians to the racial distribution of various “at-risk” civilians who could potentially be stopped. Determining the probability of an interaction is essentially a search for the correct “risk set.”

Panel A of table 1 provides a series of estimates of racial differences in the probability of police interaction by defining the relevant risk set in various ways. The first three columns use NYC Stop and Frisk data. Column 1 assumes the population at risk of being stopped by police as 18–34-year-old males. Column 2 assumes that the risk set is arrestees for 10 broadly defined felony and misdemeanor crimes as determined by

TABLE 1
RACIAL DIFFERENCES IN PROBABILITY OF INTERACTION AND OUTCOMES TEST

	NYC STOP AND FRISK			PPCS		HOUSTON		
	Population Weighted (1)	Arrest Rate Weighted (2)	Felony Arrest Weighted (3)	Involuntary Contact (4)	Voluntary Contact (5)	Population Weighted (6)	Arrest Rate Weighted (7)	Felony Arrest Weighted (8)
A. Potential selection	4.233	1.428	1.026	.922*** (.025)	.791*** (.026)	4.347	1.008	.872
B. Outcomes test	Civilian Arrested (1)	Contraband/ Weapon Found (2)		Civilian Arrested (3)	Contraband/ Weapon Found (4)			
	1.080*** (.032)	.777*** (.033)		1.814*** (.159)	.524*** (.118)			

NOTE.—This table reports the probability of police interaction and conducts outcomes tests. The sample in the NYC Stop and Frisk block consists of all NYC stop and frisks from 2003–13. The sample in the PPCS block consists of all Police-Public Contact Survey respondents from 1996–2011. The sample in the Houston block consists of all officer-involved shootings from Houston between 2000 and 2015. All population demographics have been obtained from the American Community Survey, 2007–11. Arrest rates for New York City have been obtained from NYC Enforcement Reports, 2008–13. Arrest rates for Houston have been provided from the Houston Police Department for all arrests in 2015. See the appendix for the procedures used to obtain the numbers in panel A. For panel B, we report odds ratios on the black dummy from logistic regressions of the outcome specified on data set-specific controls. For NYC Stop and Frisk, the controls are civilian demographics (race, gender, quadratic in age), encounter characteristics (stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area, during a high-crime time, or in a high-crime area at a high-crime time, whether the officer was in uniform, civilian ID type, and whether others were stopped during the interaction), civilian behavior (whether the civilian was carrying a suspicious object, if he fit a relevant description, if he was preparing for a crime, if he was on the lookout for a crime, if he was dressed in criminal attire, if there was an appearance of a drug transaction, whether there were any suspicious movements, if he was engaged in violent crime, if he was concealing a suspicious object, and whether there was any other suspicious behavior), precinct and year fixed effects, and missing indicators for all variables. For PPCS, the controls are civilian demographics (race, gender, employment last week, income, population size of a civilian's address, and a quadratic in age), contact and officer characteristics (time of day of the contact, contact type, and officer race), civilian behavior (civilian disobeyed, tried to get away, resisted, complained, argued, threatened officer, used physical force), year, and missing indicators for all variables.

* Significant at the 1 percent level.

** Significant at the 5 percent level.

*** Significant at the 10 percent level.

the New York City Police Department's Crime Reporting System. Felonies include murder and nonnegligent manslaughter, rape, other felony sex crimes, robbery, felonious assault, grand larceny, and felony crime mischief. Misdemeanor crimes include misdemeanor sex crimes, misdemeanor assault, petit larceny, and misdemeanor criminal mischief.²¹ Column 3 is similar to column 2 but includes only the six felonies.

For each of the 77 precincts, we calculate the average fraction of stops that are black and the corresponding fraction for whites. We also calculate the fraction of blacks in the relevant risk set and the same fraction for whites, for all precincts. We then regress the fraction of police stops that are black (white) on the fraction of blacks (white) in the relevant risk set and store the coefficient. The number displayed in each column is the coefficient for blacks divided by the coefficient for whites for the relevant risk set. A number greater than one indicates a potential bias against blacks. A number less than one indicates a potential bias in favor of blacks.

A simple—and often used—method to do this is to compare the fraction of blacks involved in interactions with police with their proportion in the population, though many social scientists have argued against this approach (Fridell 2004; Anwar and Fang 2006; Ridgeway 2007). Column 1 demonstrates that blacks are almost four times more likely to be stopped by police relative to their population proportion. Yet this quantity is difficult to interpret. As Fridell (2004) argues, “racial/ethnic groups are not equivalent in the nature and extent of their . . . law violating behavior” (8n1).

Column 2 uses incident weighted averages (crimes committed more often are more heavily weighted) for 10 felonies and misdemeanors. Unfortunately, we do not have racial breakdown of crime rates for individual precincts. In lieu of this, we calculate the fraction of arrestees in crimes for New York City for each year between 2008 and 2013. Conditioning on incident-weighted crime rates reduces the estimate of bias in police interactions from 4.23 to 1.43—a 66.2 percent reduction.

Column 3 conducts a similar exercise using six broad felonies. This method decreases the estimate of bias in police stopping behavior to 1.03. If one were to use robbery rates rather than all felonies, the number would be 0.546, implying that blacks are 45.4 percent less likely to be stopped (not shown in tabular form).

Column 4 in panel A of table 1 investigates potential selection into the PPCS data set. Relative to NYC's Stop and Frisk data, the PPCS data involve a larger set of police interactions and are not the result of a particular form of aggressive policing. Also, the data are from civilians' perspec-

²¹ Contents of all broad crime categories are provided in detail in any of the annual Crime and Enforcement Activity Reports released by the New York City Police Department.

tive. This allows one to analyze the probability of having an involuntary interaction with the police controlling for race and other demographics, for all respondents of the survey. In some ways, this is closer to the ideal data set described above, though we cannot control for pre-interaction civilian behavior. Involuntary interaction is a dummy variable coded to be 1 if the civilian reported that he was involved in an interaction with the police that was not initiated by him (e.g., traffic or parking violation, police asked respondent questions). The variable is coded to be 0 if the civilian reported no interaction with police or an interaction that was initiated by himself (e.g., reporting a crime, asking for assistance).

We estimate a logistic regression of involuntary interaction on civilian race, demographic variables such as gender, age, income categories, the population size of the civilian's address, a dummy variable indicating whether the civilian was employed last week or not, and year, and we report the odds ratio on the black coefficient. The odds that blacks have an involuntary interaction with police is 8 percent less than that of whites. For comparison we also provide the odds ratio for voluntary interactions. Voluntary interactions include all interactions with police that civilians initiated themselves. Blacks are 21 percent less likely to report a voluntary interaction with the police than whites.

The final three columns in panel A of table 1 report estimates from an analysis identical to the one conducted for the Stop and Frisk data set, but for Houston officer-involved shootings.²² Column 6 demonstrates that blacks are 4.35 times more likely to be involved in an officer-involved shooting than nonblacks relative to their proportion in the 18–34-year-old male population. This estimate changes drastically to 1.01—a 76.8 percent reduction—when the population defined at risk is the fraction of arrestees in felonies and misdemeanors. The estimate decreases further to 0.87 when only felony crimes are taken into account.

Panel B of table 1 reports the results of a series of Beckerian outcomes tests (Becker 1993), where the outcomes are whether or not a police stop resulted in an arrest or whether contraband or any weapon was found. Becker, in the context of mortgages, argued that discrimination in mortgage lending against blacks cannot be found simply by looking at the likelihood of getting a loan for minority versus white applicants who are similar in incomes, credit backgrounds, and other available characteristics. The correct procedure would be to determine whether loans are more profitable to blacks (and other minorities) than to whites. Discriminating banks would turn down marginally profitable black applicants but accept white applicants. This is the spirit behind the seminal work in Knowles et al. (2001).

²² Potential selection into all OIS locations by population weights and UCR coded arrest rates are presented at the end of app. table 2C.

For the outcomes test, we estimate a logistic regression of whether the civilian was arrested/was carrying contraband or weapons on race, civilian demographics, encounter characteristics, civilian behavior, and suitable fixed effects.²³ We report the odds ratio on the black coefficient. If the coefficient is above one, this implies that stops of blacks are more “productive” than those of whites, and thus, if anything, police should be stopping blacks more at the margin.

Unfortunately, whether or not there seems to be racial bias in police stopping behavior depends on the outcome tested. When using whether or not the civilian was arrested as an outcome—which has the important disadvantage of depending on the subsequent behavior of both civilians and police—there seems to be no bias against blacks in police stopping behavior. In other words, blacks are more likely to be arrested, conditional on being stopped. When the outcome is whether or not contraband or a weapon was found, black stops are significantly less productive than those of whites and thus are evidence for potential bias.

Taken together, this evidence demonstrates how difficult it is to understand whether there is potential selection into police data sets. Estimates range from blacks being 323 percent more likely to be stopped to 47.6 percent less likely to be stopped. Solving this is outside the scope of this paper, but the data suggest the following rough rule of thumb: if one assumes that police are nonstrategic in stopping behavior, there is bias. Conversely, if one assumes that police are stopping individuals they are worried will engage in violent crimes, the evidence for bias is exceedingly small.

IV. Estimating Racial Differences in Nonlethal Use of Force

A. NYC’s Stop, Question, and Frisk Data

Table 2 presents a series of estimates of racial differences in police use of force, conditional on being stopped, using the Stop and Frisk data. We estimate logistic regressions of the following form:

$$\ln \left(\frac{\Pr(\text{Force}_{i,p,t} = 1)}{1 - \Pr(\text{Force}_{i,p,t} = 1)} \right) = \text{Race}'_i \alpha + X'_{i,t} \beta + Z'_{p,t} \mu + \nu_t + \psi_p + \varepsilon_{i,p,t}, \quad (1)$$

where $\text{Force}_{i,p,t}$ is a measure of police use of force on individual i , in precinct p , at time t . A full set of race dummies for civilians are included in the regressions, with white as the omitted category. Consequently, the coefficients on race capture the gap between the named racial category

²³ All controls used are reported in detail in summary statistics in app. tables 2A and 2B.

TABLE 2
RACIAL DIFFERENCES IN NONLETHAL USE OF FORCE, CONDITIONAL ON AN INTERACTION

	White Mean (1)	Black (2)	Hispanic (3)	Asian (4)	Other Race (5)
A. NYC Stop, Question, and Frisk (<i>N</i> = 4,927,962)					
a. No controls	.153	1.534*** (.144)	1.582*** (.149)	1.044 (.119)	1.392*** (.121)
b. + civilian demographics		1.480*** (.146)	1.517*** (.146)	1.010 (.122)	1.346*** (.114)
c. + encounter characteristics		1.655*** (.155)	1.641*** (.157)	1.059 (.133)	1.452*** (.121)
d. + civilian behavior		1.462*** (.128)	1.516*** (.136)	1.051 (.124)	1.372*** (.107)
e. + precinct fixed effects, year fixed effects		1.178*** (.034)	1.122*** (.026)	.953 (.033)	1.060** (.028)
f. + precinct × year fixed effects		1.171*** (.034)	1.112*** (.025)	.954 (.033)	1.066** (.028)
g. + precinct × year × month fixed effects		1.172*** (.034)	1.112*** (.025)	.958 (.032)	1.068** (.028)
B. Police-Public Contact Survey (<i>N</i> = 59,668)					
h. No controls	.007	3.496*** (.364)	2.697*** (.311)	...	1.130 (.275)
i. + civilian demographics		2.745*** (.299)	1.716*** (.205)792 (.195)
j. + encounter characteristics		2.659*** (.293)	1.695*** (.202)811 (.197)
k. + civilian behavior		2.780*** (.330)	1.820*** (.225)763 (.194)
l. + year		2.769*** (.328)	1.818*** (.225)758 (.193)

NOTE.—This table reports odds ratios obtained from logistic regressions. The sample in panel A consists of all NYC stop and frisks from 2003–13 with nonmissing use of force data. The dependent variable is an indicator for whether the police reported using any force during a stop and frisk interaction. The omitted race is white, and the omitted ID type is other. Col. 1 gives the unconditional average of stop and frisk interactions that reported any force being used for white civilians. Cols. 2–5 report logistic estimates for black, Hispanic, Asian, and other race civilians, respectively. Each row corresponds to a different empirical specification. Row a includes solely racial group dummies. Row b adds controls for gender and a quadratic in age. Row c adds controls for whether the stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area, during a high-crime time, or in a high-crime area at a high-crime time, whether the officer was in uniform, civilian ID type, and whether others were stopped during the interaction. Row d adds controls for civilian behavior. Row e adds precinct and year fixed effects. Row f adds precinct×year fixed effects. Row g adds precinct×year×month fixed effects. Each row includes missing indicators in all variables. Standard errors, clustered at the precinct level, are reported in parentheses. The sample in panel B consists of all PPCS respondents from 1996–2011 with nonmissing use of force data. The dependent variable is an indicator for whether the survey respondent reported any force being used in a contact with the police. The omitted race is white. Col. 1 gives the unconditional average of contacts in which survey respondents reported any force being used for white civilians. Cols. 2–5 report logistic estimates for black, Hispanic, and other race civilians, respectively. Each row corresponds to a different empirical specification. Row h includes solely racial group dummies. Row i adds controls for civilian gender, work, income, population size of a civilian's address, and a quadratic in age. Row j adds controls for the time of day of the contact, contact type, and officer race. Row k adds an indicator for civilian behavior. Row l adds a control for year. Each row includes missing indicators in all variables. Standard errors, robust to heteroskedasticity, are reported in parentheses.

* Significant at the 1 percent level.

** Significant at the 5 percent level.

*** Significant at the 10 percent level.

and whites, which is reported as an odds ratio.²⁴ The vectors of covariates included in the specification, denoted $X_{i,t}$ and $Z_{p,t}$, vary between rows in table 2. As one moves down the table, the set of coefficients steadily grows. We caution against a causal interpretation of the coefficients on the covariates, which are better viewed as proxies for a broad set of environmental and behavioral factors at the time of an incident. Standard errors, which appear below each estimate, are clustered at the precinct level unless otherwise specified.

Row a in table 2 presents the differences in means for any use of force conditional on a police interaction. These results reflect the raw gaps in whether or not a police stop results in any use of force, by race. Blacks are 53 percent more likely to experience any use of force relative to a white mean of 15.3 percent. The raw gap for Hispanics is almost identical. Asians are no more likely than whites to experience use of force. The raw gap for other race—which included American Indians, Alaskan natives, or other races besides white, black, Hispanic, and Asian—is smaller but still considerable.

The raw difference between races is large—perhaps too large—and it seems clear that one needs to account for at least some contextual factors at the time of a stop in order to better understand, for example, whether racial differences are driven by police response to a given civilian's behavior or racial differences in civilian behavior. Yet, it is unclear how to account for context that might predict how much force is used by police and not include variables that themselves might be influenced by biased police.²⁵

Row b adds baseline civilian characteristics—such as age and gender—all of which are exogenously determined and not strategically chosen as a function of the police interaction. Adding these variables does almost nothing to alter the odds ratios. Encounter characteristics—whether the interaction happened inside, the time of day, whether it occurred in a high- or low-crime area, and whether the civilian provided identification—are added as controls in row c. If anything, adding these variables increases the odds ratios on each race, relative to whites. Surprisingly, accounting for civilian behavior—row d in the table—does little to alter the results.

Row e in table 2 includes both precinct and year fixed effects. This significantly changes the magnitude of the coefficients. Blacks are almost 18 percent more likely to incur any use of force in an interaction, ac-

²⁴ Appendix tables 3A–3G run similar specifications using ordinary least squares and obtain similar results. Estimating probit models provides almost identical results.

²⁵ The traditional literature in labor economics—beginning with Mincer (1958)—dealt with similar issues. O'Neill (1990) and Neal and Johnson (1996) sidestep this by demonstrating that much of the racial wage gap can be accounted for by including only premarket factors such as test scores.

counting for all variables we can in the data. Hispanics are roughly 12 percent more likely.²⁶ Both are statistically significant. Asians are slightly less likely, though not distinguishable from whites. Row f interacts precincts with year as fixed effects. Results do not change significantly from row e. Changing fixed effects to be interactions between precinct, year, and month (row g) does not alter the results.

These data have two potential takeaways: precincts matter, and accounting for a large and diverse set of control variables, black civilians are still more likely to experience police use of force. Of the 112 variables available in the data, there is no linear combination that fully explains the race coefficients.²⁷ From this point forward, we consider the row e specification, including precinct and year fixed effects as our main specification.

Inferring racial differences in the types of force used in a given interaction is a bit more nuanced. Police report that in 20 percent of all stops, some use of force is deployed. Officers routinely record more than one use of force. For instance, a stop might result in an officer putting his hands on a civilian, who then pushes the officer, and the officer responds by pushing him to the ground. This would be recorded as “hands” and “force to ground.” In 85.1 percent of cases, exactly one use of force is recorded. Two use of force categories were used in 12.6 percent of cases, 1.8 percent report three use of force categories, and 0.6 percent of all stop and frisk incidents in which force is used record more than three uses of force.

There are several ways to handle this. The simplest is to code the max force used as 1 and all the lower-level uses of force in that interaction as 0. In the example above in which an officer recorded both “hands” and “forced to the ground” as uses of force, one would ignore the use of hands and code forced to the ground as 1. The limitation of this approach is that it discards potentially valuable information on lower-level uses of force. When analyzing racial differences in the use of hands by police, one would miss this observation. A similar issue arises if one uses the parallel “min.”²⁸

Perhaps a more intuitive way to code the data is to treat each use of force as “at least as much.” In the example above, both hands and forced

²⁶ Even accounting for eventual outcomes of each stop—which include being let go, being frisked, being searched, being arrested, being summonsed, and whether or not a weapon or some form of contraband was found—blacks are 22 percent more likely to experience force and Hispanics are 27 percent more likely. We did not include these control variables in our main specification because of the fear of overcontrolling if there is discrimination in the probability of arrests, conditional on behavior.

²⁷ Using data on geospatial coordinates, we also included block-level fixed effects, and the results were qualitatively unchanged.

²⁸ Appendix tables 9A–9C demonstrate that altering the definition to be “at most” or using the max/min force used in any given police interaction does not alter the results.

to the ground would be coded as 1 in the raw data. When analyzing racial differences in the use of hands by police, this observation would be included. The interpretation would not be racial differences in the use of hands, *per se*, but racial differences in the use of “at least” hands. To be clear, an observation that records only hands would be in the hands regression but not in the regression that restricts the sample to observations in which individuals were at least forced to the ground. This is the method we use throughout.

Results using this method to describe racial differences for each use of force are displayed in figure 1. The *x*-axis contains use of force variables that range from at least hands to at least the use of pepper spray or a baton. The *y*-axis measures the odds ratio for blacks (panel A) or Hispanics (panel B). The solid black line is gleaned from regressions with no controls, and the dashed gray line adds all controls and precinct and year fixed effects (equivalent to row e in table 2).

For blacks, the consistency of the odds ratios is striking. As the use of force increases, the frequency with which that level of force is used decreases substantially. There are approximately 5 million observations in the data: 19 percent of them involve the use of hands while 0.04 percent involve using pepper spray or a baton. The use of high levels of force in these data is rare. Yet it is consistently rarer for whites relative to blacks. Odds ratios across all levels of force range between 1.175 (0.036) and 1.275 (0.131).

Interestingly, for Hispanics, once we account for our set of controls, there are small differences in use of force for the lower-level uses of non-lethal force, but the differences converge toward whites as the use of force increases both in the raw data and with the inclusion of controls.

One may be concerned that restricting all the coefficient estimates to be identical across the entire sample may yield misleading results. Regressions on a common support (e.g., only on males or only on police stops during the day) provide one means of addressing this concern. Panel A of table 3 explores the sensitivity of the estimated racial gaps in police use of force across a variety of subsamples of the data. I report only the odds ratios on black and Hispanic and associated standard errors. The top row of the table presents baseline results using the full (any force) sample and our parsimonious set of controls (corresponding to row e in table 2). The subsequent rows investigate racial differences in use of force for high-/low-crime areas, time of day, whether or not the officer was in uniform, indoors/outdoors, gender of civilian, and eventual outcomes.

Most of the coefficients on race do not differ significantly at the 1 percent level across these various subsamples with the exception of time of day and eventual outcomes. Black civilians are 8.6 percent more likely to have any force used against them conditional on being arrested. They

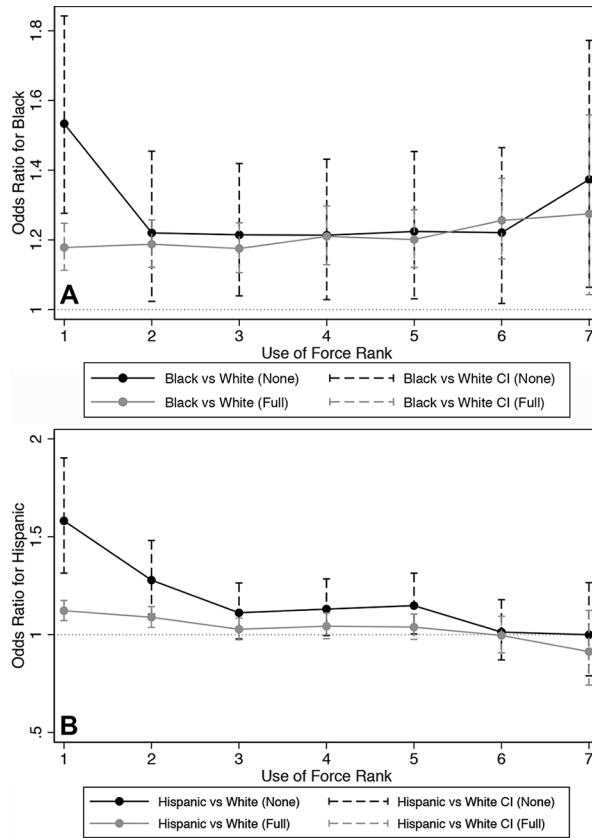


FIG. 1.—Odds ratios by use of force (conditional on an interaction), NYC Stop, Question, and Frisk. These figures plot odds ratios with 95 percent confidence intervals from logistic regressions. For panel A, the y-axis denotes the odds ratio of reporting various uses of force for black civilians versus white civilians. For panel B, the y-axis denotes the odds ratio of reporting various uses of force for Hispanic civilians versus white civilians. For both panels, the x-axis denotes different use of force types: 1 is an indicator for whether the police reported using at least hands or more severe force on a civilian in a stop and frisk interaction; 2 is for whether the police reported at least pushing a civilian to a wall or using more severe force; 3 is for whether the police reported at least using handcuffs or more severe force; 4 is for whether the police reported at least drawing a weapon on a civilian or using more severe force; 5 is for whether the police reported at least pushing a civilian to the ground or using more severe force; 6 is for whether the police reported at least pointing a weapon at a civilian or using more severe force; finally, 7 is for whether the police reported at least using pepper spray or a baton on a civilian. All force indicators are coded as 0 when the police report using no force in a stop and frisk interaction. The line plot with no controls is achieved by regressing the type of force (described above) on civilian race dummies only. The line plot with full controls is achieved by regressing the type of force on civilian race dummies, civilian gender, a quadratic in age, civilian behavior, whether the stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area or at a high-crime time, whether the officer was in uniform, civilian ID type, whether others were stopped during the interaction, and missing indicators in all variables. Precinct and year fixed effects were included in the controlled regression. Standard errors are clustered at the precinct level.

are 15.6 percent more likely to have any force used against them conditional on being summonsed and 12.7 percent more likely conditional on having weapons or contraband found on them. Results are similar for Hispanics. Additionally, for both blacks and Hispanics, racial differences in use of force are more pronounced during the day relative to night.

To dig deeper, panel A in figure 2 plots the odds ratios of any use of force for black civilians versus white civilians for every hour of the day. Panel B displays the average use of force for black civilians and white civilians for every hour of the day. These figures show that force against black civilians follows approximately the same pattern as against white civilians, though the difference between average force between the two races decreases at night.

B. Police-Public Contact Survey

One of the key limitations of the Stop and Frisk data is that one gets only the police side of the story or, more accurately, the police entry of the data. It is plausible that there are large racial differences that exist that are masked by police misreporting. The PPCS is one way to partially address this weakness.

Table 2, panel B, presents a series of estimates of racial differences in police use of force conditional on an interaction, using the PPCS data. The specifications estimated have the following form:

$$\ln\left(\frac{\Pr(\text{Force}_{i,t} = 1)}{1 - \Pr(\text{Force}_{i,t} = 1)}\right) = \text{Race}'_i\alpha + X'_{i,t}\beta + \nu_t + \varepsilon_{i,t},$$

where $\text{Force}_{i,t}$ is a measure of police use of force reported by individual i in year t . A full set of race dummies for individuals and officers are included in the regressions, with white as the omitted category. The vectors of covariates included in the specification vary across rows in table 2, panel B. As one moves down the table, the set of coefficients steadily grows. Standard errors, which appear below each estimate, account for heteroskedasticity.

Generally, the data are qualitatively similar to the results using Stop and Frisk; namely, despite a large and complex set of controls, blacks and Hispanics are more likely to experience some use of force from police. A key difference, however, is that the share of individuals experiencing any use of force is significantly lower. In the Stop and Frisk data, 15.3 percent of whites incur some force in a police interaction. In the PPCS, this number is 1 percent. There are a variety of potential reasons for these stark differences. For instance, the PPCS is a nationally representative sample of interactions with police from across the United States, whereas the Stop and Frisk data are gleaned from a rather aggressive proactive policing

TABLE 3
ANALYSIS OF SUBSAMPLES, ANY USE OF FORCE, CONDITIONAL ON AN INTERACTION

	White Mean (1)	Coefficient on Black (2)	Coefficient on Hispanic (3)	Observations (4)
Full sample	.153	1.178*** (.034)	1.122*** (.026)	4,927,962
A. NYC Stop, Question, and Frisk				
1. Crime rate in area:				
High crime	.143	1.170*** (.035)	1.118*** (.027)	2,750,559
Low crime	.163	1.202*** (.039)	1.139*** (.029)	2,177,403
<i>p</i> -value		.254	.320	
2. Time of day:				
Day	.126	1.260*** (.035)	1.164*** (.026)	1,783,977
Night	.170	1.141*** (.039)	1.102*** (.029)	3,141,371
<i>p</i> -value		.001	.024	
3. Officer in uniform:				
Uniformed officer	.132	1.180*** (.047)	1.126*** (.035)	3,546,388
Nonuniformed officer	.189	1.200*** (.033)	1.124*** (.023)	1,381,074
<i>p</i> -value		.717	.954	
4. Location:				
Indoors	.144	1.143*** (.044)	1.105*** (.033)	1,129,555
Outdoors	.154	1.186*** (.031)	1.125*** (.025)	3,771,939
<i>p</i> -value		.241	.504	
5. Civilian gender				
Male	.160	1.175*** (.034)	1.122*** (.026)	4,447,382
Female	.089	1.255*** (.055)	1.109*** (.043)	343,199
<i>p</i> -value		.042	.717	
6. Eventual outcomes:				
Frisked	.312	1.036 (.024)	1.022 (.021)	2,725,795
Searched	.412	1.061* (.038)	1.043 (.031)	415,455
Arrested	.327	1.086*** (.035)	1.045* (.025)	291,166
Summonsed	.195	1.156*** (.044)	1.068** (.035)	304,603
Weapon/contraband found	.359	1.127*** (.026)	1.068*** (.024)	136,926
<i>p</i> -value		.002	.339	

TABLE 3 (Continued)

	White Mean (1)	Coefficient on Black (2)	Coefficient on Hispanic (3)	Observations (4)
B. Police-Public Contact Survey				
Full sample	.007	2.769*** (.328)	1.818*** (.225)	59,668
1. Officer race:				
Black/Hispanic	.005	2.089 (1.336)	5.584*** (3.048)	2,166
White	.008	2.823*** (.556)	1.883*** (.401)	21,456
<i>p</i> -value		.653	.064	
2. Civilian gender:				
Male	.011	2.827*** (.384)	1.912*** (.258)	30,154
Female	.003	2.588*** (.747)	1.433 (.377)	28,835
<i>p</i> -value				
3. Time of day:				
Daytime	.004	3.690*** (.976)	2.368*** (.614)	16,324
Nighttime	.012	1.848** (.520)	2.332*** (.608)	7,640
<i>p</i> -value		.073	.966	
4. Civilian income:				
\$0–\$20,000	.010	2.944*** (.534)	1.630** (.334)	15,014
\$20,000–\$50,000	.008	2.010*** (.491)	1.890*** (.420)	14,314
\$50,000+	.004	3.942*** (1.273)	1.761 (.680)	19,246
<i>p</i> -value		.220	.887	

NOTE.—This table reports odds ratios obtained from logistic regressions. For panel A, the sample consists of all NYC stop and frisks from 2003–13 in which use of force and reported subgroup variables were nonmissing. The dependent variable is whether any force was used during a stop and frisk interaction, with each row presenting results from indicated subgroups. We control for gender, a quadratic in age, civilian behavior, whether the stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area, during a high-crime time, or in a high-crime area at a high-crime time, whether the officer was in uniform, civilian ID type, whether others were stopped during the interaction, and missing indicators in all variables. Precinct and year fixed effects were included in all regressions. For panel B, the sample consists of all PPCS respondents from 1996–2011 in which use of force and reported subgroup variables were nonmissing. The dependent variable is whether any force was used during a contact, with each panel presenting results from indicated subgroups. We control for civilian gender, a quadratic in age, work, income, population size of a civilian's address, civilian behavior, contact time, contact type, officer race, year of survey, and missing indicators in all variables. Standard errors, clustered at the precinct level, are reported in parentheses.

* Significant at the 1 percent level.

** Significant at the 5 percent level.

*** Significant at the 10 percent level.

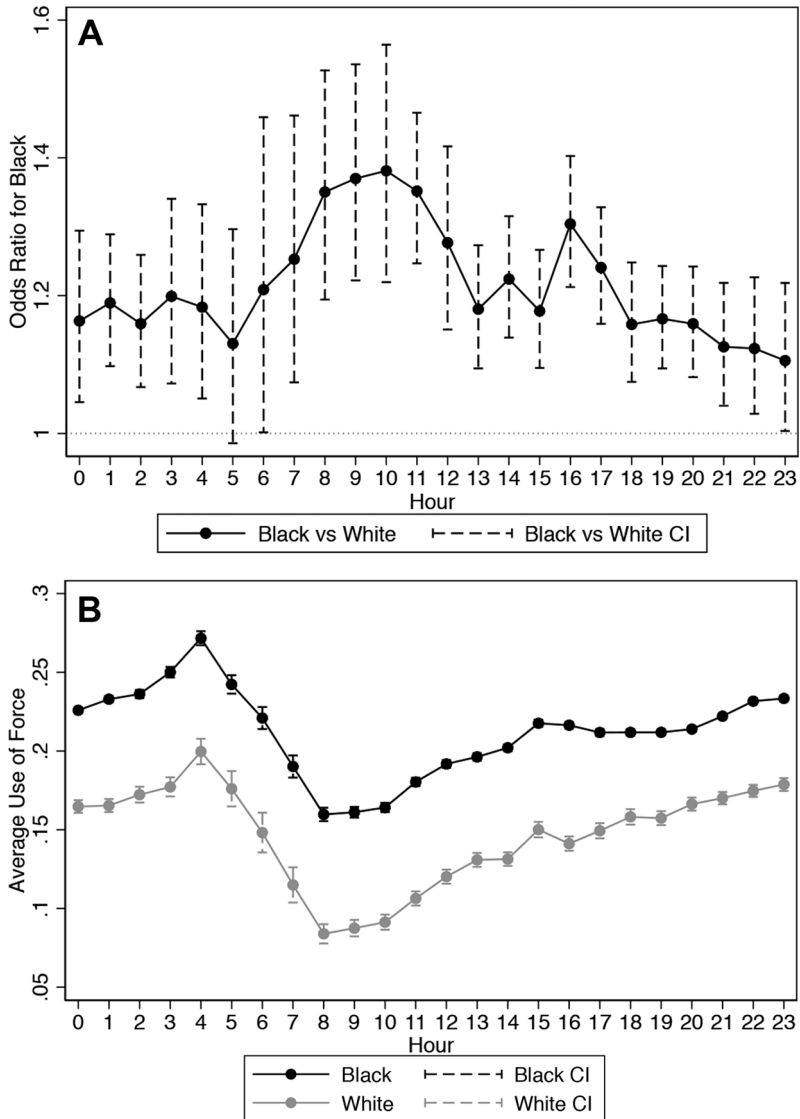


FIG. 2.—Odds ratios of any use of force (conditional on an interaction) by time of day, NYC Stop, Question, and Frisk. These figures plot odds ratios with 95 percent confidence intervals and averages with 95 percent confidence intervals. For panel A, the y-axis denotes the odds ratio of reporting any use of force for black civilians versus white civilians. For panel B, the y-axis denotes the average fraction of white and black civilians who had any force used against them. For both figures, the x-axis denotes different hours of the day. For panel A, odds ratios are achieved by regressing any use of force on civilian race dummies, civilian gender, a quadratic in age, civilian behavior, whether the stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area or at a high-crime time, whether the officer was in uniform, civilian ID type, whether others were stopped during the interaction, and missing indicators in all variables, for every hour of the day. Precinct and year fixed effects were included in all regressions. Standard errors are clustered at the precinct level.

strategy in a large urban city. This is important because in what follows we present odds ratios. Odds ratios are informative, but it is important for the reader to know that the baseline rate of force is substantially smaller in the PPCS.

Blacks are 3.5 times more likely to report use of force by police in an interaction in the raw data. Hispanics are 2.7 times more likely. Adding controls for demographic and encounter characteristics, civilian behavior, and year reduces the odds ratio to roughly 2.8 for blacks and 1.8 for Hispanics. Differences in quantitative magnitudes aside, the PPCS paints a similar portrait: large racial differences in police use of force that cannot be explained using a large and varied set of controls.

One important difference between the PPCS and the Stop and Frisk data concerns racial differences on the more extreme uses of nonlethal force: using pepper spray or striking with a baton. Recall that in the Stop and Frisk data, the odds ratios were relatively consistent as the intensity of force increased. In the PPCS data, if anything, racial differences on these higher uses of force disappear. For kicking or using a stun gun or pepper spray, the highest use of force available, the black coefficient is 1.930 (0.649) and the Hispanic coefficient is 1.446 (0.490), though because of the rarity of these cases the coefficients are barely statistically significant at the 5 percent level (see fig. 3).

Panel B of table 3 explores the heterogeneity in the data by estimating racial differences in police use of force in the PPCS on various subsamples of the data: officer race, civilian income, civilian gender, and time of contact. Civilian income is divided into three categories: less than \$20,000, between \$20,000 and \$50,000, and above \$50,000. Strikingly, both the black and Hispanic coefficients are statistically similar across these income levels—suggesting that higher-income minorities do not price themselves out of police use of force—echoing some of the ideas in Cose (1993). Racial differences in police use of force do not seem to vary with civilian gender or officer race especially for black civilians. Consistent with the results in the Stop and Frisk data, the black coefficient is 3.690 (0.976) for interactions that occur during the day and 1.848 (0.520) for interactions that occur at night. The *p*-value on the difference is significant, but only at the 10 percent level.

Putting the results from the Stop and Frisk and PPCS data sets together, a pattern emerges. Relative to whites, blacks and Hispanics seem to have very different interactions with law enforcement—interactions that are consistent with, though definitely not proof of, some form of discrimination. Including myriad controls designed to account for civilian demographics, encounter characteristics, civilian behavior, eventual outcomes of interaction, and year reduces, but cannot eliminate, racial differences in nonlethal use of force in either of the data sets analyzed.

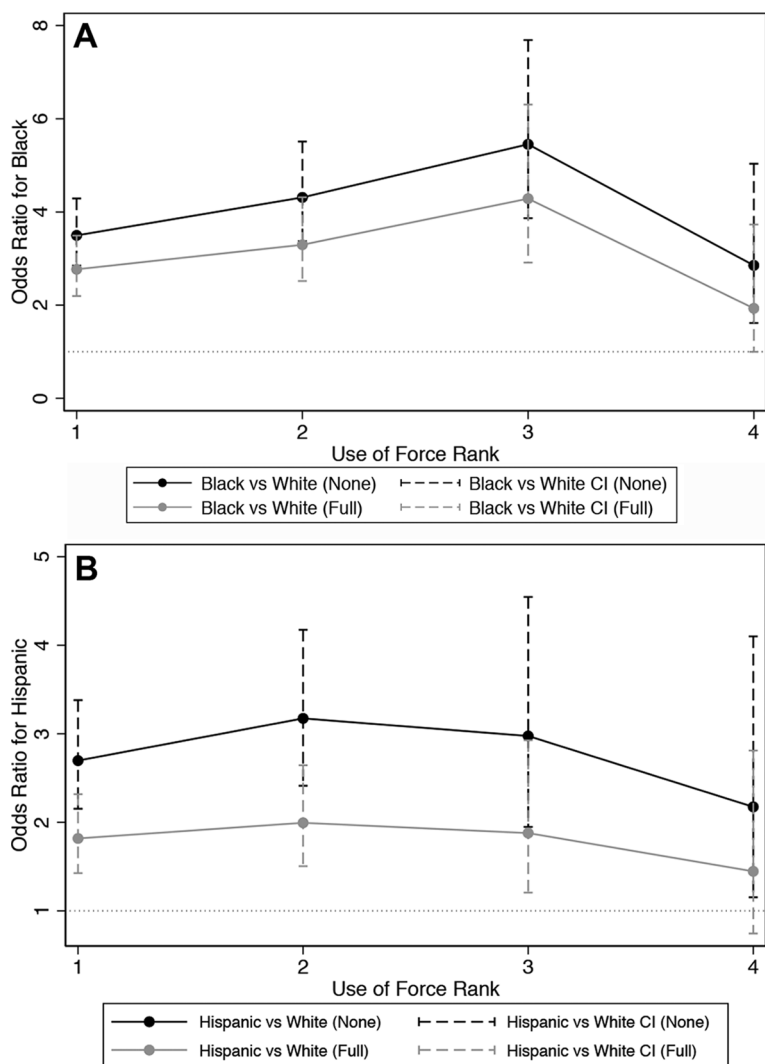


FIG. 3.—Odds ratios by use of force (conditional on an interaction), Police-Public Contact Survey. These figures plot odds ratios with 95 percent confidence intervals from logistic regressions. For panel A, the y-axis denotes the odds ratio of reporting various uses of force for black civilians versus white civilians. For panel B, the y-axis denotes the odds ratio of reporting various uses of force for Hispanic civilians versus white civilians. For both figures, the x-axis denotes different use of force types: 1 is an indicator for whether the survey respondent reported the officer at least grabbing him or her in an interaction; 2 is for whether the respondent reported the police handcuffing him or her or using more severe force in an interaction; 3 is for whether the survey respondent reported the police pointing a gun at him or her or using more severe force in an interaction; and, finally, 4 is for whether the respondent reported the police kicking, using a stun gun, or using pepper spray on him or her or using more severe force. All force indicators are coded as 0 when the respondent reports the police using no force in an interaction. The line plot with no controls is achieved by regressing the type of force (described above) on civilian race dummies only. We control for civilian gender, a quadratic in age, work, income, population size of the civilian's address, civilian behavior, contact time, contact type, officer race, year of survey, and missing indicators in all variables. Standard errors are robust.

V. Estimating Racial Differences in Officer-Involved Shootings

We now focus on racial differences in officer-involved shootings. We begin with specifications most comparable to those used to estimate racial differences in nonlethal force, using both data from officer-involved shootings in Houston and data we coded from Houston arrest records that contain interactions with police that might have resulted in the use of lethal force.²⁹ Specifically, we estimate the following empirical model:

$$\ln\left(\frac{\Pr(\text{shooting}_{i,t})}{1 - \Pr(\text{shooting}_{i,t})}\right) = \text{Race}'_i\alpha + X'_{i,t}\beta + \nu_t + \varepsilon_{i,t},$$

where $\text{shooting}_{i,t}$ is a dichotomous variable equal to one if a police officer discharged his weapon at individual i in year t . There are no accidental discharges in our data, and shootings at canines have been omitted. A full set of race dummies for individuals and officers are included in the regressions, with nonblack, non-Hispanics as the omitted category for individuals. The vectors of covariates included in the specification vary across rows in table 4. As one moves down the table, the set of coefficients generally grows. As one moves across the columns of the table, the comparison risk set changes.³⁰ Presenting the results in this way is meant to underscore the robustness of the results to the inclusion of richer sets of controls and to alternative interpretations of the risk sets. Standard errors, which appear below each estimate, account for heteroskedasticity.

Given the stream of video “evidence,” which many take to be indicative of structural racism in police departments across America, the ensuing and understandable outrage in black communities across America, and the results from our previous analysis of nonlethal uses of force, the results displayed in table 4 are startling.

Blacks are 23.5 percent less likely to be shot by police, relative to whites, in an interaction. Hispanics are 8.5 percent less likely to be shot, but the coefficient is statistically insignificant.

Rows b–f add various controls, identical to those in appendix table 2C. Accounting for basic suspect or officer demographics does not significantly alter the raw racial differences. Including encounter characteristics—which one can accomplish only by hand-coding the narratives embedded in arrest reports—creates more parity between blacks and nonblack, non-Hispanic suspects, rendering the coefficient closer to one. Finally,

²⁹ Because of this select set of zeros, the nonblack, non-Hispanic mean, displayed in col. 1 of table 4, is drastically larger than a representative sample of the population—which would be approximately 0.0001 percent. In our data, 45.5 percent of whites were involved in an officer-involved shooting.

³⁰ Appendix table 7 investigates the sensitivity of the main results to more alternative compositions of the risk sets.

TABLE 4
RACIAL DIFFERENCES IN LETHAL USE OF FORCE, CONDITIONAL ON AN INTERACTION, EXTENSIVE MARGIN, OFFICER-INVOLVED SHOOTINGS

	APPROXIMATE OIS WITH NARRATIVES			TASER WITHOUT NARRATIVES		FULL SAMPLE WITHOUT NARRATIVES	
	Nonblack/Non-Hispanic (1)	Black (2)	Hispanic (3)	Nonblack Mean (4)	Black (5)	Nonblack Mean (6)	Black (7)
a. No controls	.455	.765 (.138)	.915 (.176)	.185	.636*** (.063)	.151	.673*** (.065)
b. + suspect demographics		.786 (.151)	.969 (.176)		.650*** (.066)		.683*** (.067)
c. + officer demographics		.780 (.192)	1.115 (.294)		.726** (.094)		.749*** (.087)
d. + encounter characteristics		.890 (.252)	.991 (.295)		.687*** (.098)		.754*** (.097)
e. + suspect weapon		.806 (.284)	1.333 (.489)		. . . (.099)		. . .
f. + year		.726 (.257)	1.211 (.457)		.693*** (.099)		.758*** (.098)
Observations		1,532			5,012		5,994

NOTES.—This table reports odds ratios from logistic regressions. The sample for each regression is displayed in the top row. For cols. 1–3, the sample consists of all officer-involved shootings in Houston from 2000–2015, plus a random draw of all arrests for the following offenses, from 2000–2015: aggravated assault on a peace officer, attempted capital murder of a peace officer, resisting arrest, evading arrest, and interfering in an arrest. These arrests contain narratives from police reports. For cols. 4 and 5, the sample consists of all officer-involved shootings in Houston from 2000–2015, plus a sample of

arrests in which tasers were used. These arrests do not contain narratives from police reports. For cols. 6 and 7, the sample combines all officer-involved shootings in Houston from 2000–2015, plus a random draw of all arrests for the following offenses, from 2000–2015: aggravated assault on a peace officer, attempted capital murder of a peace officer, resisting arrest, evading arrest, and interfering in an arrest, plus arrests in which tasers were used. These arrests do not contain narratives from police reports. Data without narratives have no information on officer duty, civilian's attack on officer, and civilian weapon. The dependent variable is whether the officer fired his gun during the encounter. The omitted race is nonblacks (with the exception of the sample with narratives, where the omitted race is nonblack/non-Hispanic). The first column for each sample gives the unconditional average of contacts that resulted in an officer firing his gun. The second column for each sample reports logistic estimates for black civilians. Each row corresponds to a different empirical specification. Row a includes solely racial dummies. Row b adds civilian gender and a quadratic in age. Row c adds controls for the split of races of officers present at the scene, whether any female officers were present, whether officers were on duty or not, whether multiple officers were present, and the average tenure of officers at the scene. Row d adds controls for the reason the officers were responding at the scene, whether the encounter happened during daytime, and whether the civilian attacked or drew a weapon. Row e adds controls for the type of weapon the civilian was carrying. Row f adds year fixed effects for cols. 1 and 2. It adds year as a categorical variable for cols. 4–7. Each row includes missing indicators in all variables. For arrest data without narratives, missing indicators for officer gender, officer tenure, and number of officers on the scene were removed to minimize the loss of observations in logistic regressions. For all regressions, missing indicators for response reason and for whether the civilian attacked or drew a weapon were removed for the same reason. Standard errors are robust and are reported in parentheses.

* Significant at the 1 percent level.

** Significant at the 5 percent level.

*** Significant at the 10 percent level.

when we include whether or not a suspect was found with a weapon or year fixed effects, the coefficients still suggest that, if anything, officers are less likely to shoot black suspects, *ceteris paribus*, though the racial differences are not significant.

Columns 4 and 5 of table 4 include as a risk set 4,504 incident-suspect observations from 2000–2015 for all arrests during which an officer reported using his taser, in addition to all OIS in Houston from that time period. The empirical question here is whether or not there are racial differences in the split-second decision whether to use lethal or nonlethal force through the decision to shoot a pistol or use a taser.

Consistent with the previous results, the raw racial difference in the decision to employ lethal force using this taser sample is negative and statistically significant. Adding suspect and officer demographics, encounter characteristics, and year controls does little to change the odds ratios for black versus nonblack suspects. Including all controls available from the taser sample, table 4 shows that black civilians are 30.7 percent less likely to be shot with a pistol (rather than a taser) relative to nonblack suspects. Columns 6 and 7 pool the sample from hand-coded arrest data and taser data. Results remain qualitatively the same. Controlling for all characteristics from incident reports, black suspects are 24.2 percent less likely to be shot than nonblack suspects.

To be clear, the empirical thought experiment here is that a police officer arrives at a scene and decides whether or not to use lethal force. Our estimates suggest that this decision is not correlated with the race of the suspect. This does not, however, rule out the possibility that there are important racial differences in whether or not these police-civilian interactions occur at all.

Appendix tables 6 and 7 explore the sensitivity of the results for various subsamples of the data: whether the unit that responded was majority black or Hispanic or majority white or Asian, the number of officers who responded to the scene, whether the suspect clearly drew a weapon versus appeared to draw a weapon, whether the officer was on duty, and the type of call the officer was responding to (a partial test of the selection issue described above). Equations identical to (3) are estimated, but because of the smaller sample sizes inherent in splitting the sample, we estimate ordinary least squares regressions.

None of the subsamples explored demonstrate much difference of note. We find no differences in the use of lethal force across different call slips—the *p*-value for the equality of race coefficient across different call slips is .763 for black suspects—suggesting that officers seeking confrontation in random street interactions in a way that causes important selection bias into our sample is not statistically relevant. Subsampling on the number and racial composition of the officer unit also shows no evidence of racial differences.

Another way to investigate the robustness of our coefficients is to analyze the odds ratios across time. These data are displayed in figure 4. Racial differences in OIS between 2000 and 2015 are remarkably constant. This interval is interesting and potentially informative as it is 9 years after the public beatings of Rodney King and includes the invention of Facebook, the iPhone, YouTube, and related technology that allows bystanders to capture police-civilian interactions and make it publicly available at low costs. Crudely, one might think the period between 2000 and 2005 to be years in which police misconduct could more easily go unnoticed and for which public attention was relatively low. Thus, the disincentive to misreport was likely lower. After this period, misreporting costs likely increased. Yet, as we see from figure 4, this does not seem to influence racial differences in the use of lethal force.

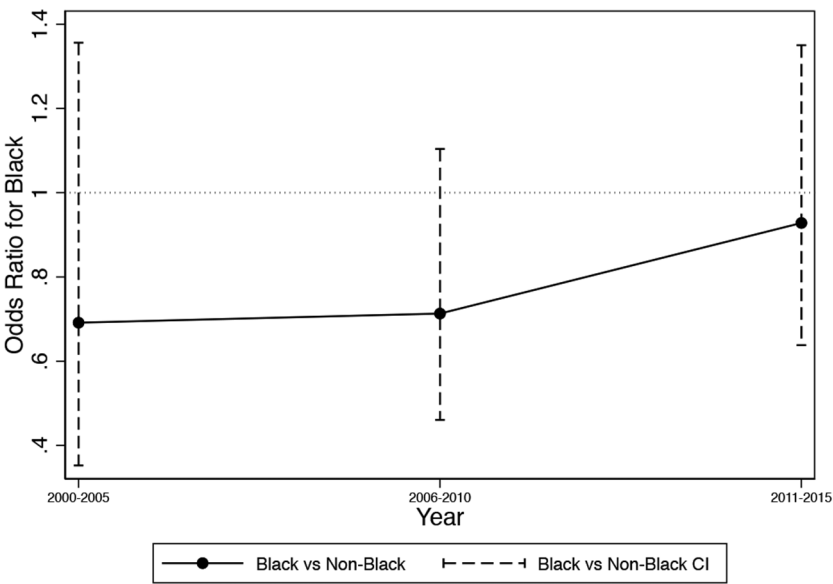


FIG. 4.—Odds ratios for officer-involved shootings (conditional on an interaction), extensive margin, by year categories. This figure plot odds ratios with 95 percent confidence intervals from logistic regressions. The sample consists of all officer-involved shootings in Houston from 2000 to 2015, plus a random draw of all arrests for the following offenses, from 2000 to 2015: aggravated assault on a peace officer, attempted capital murder of a peace officer, resisting arrest, evading arrest, and interfering in an arrest, plus a sample of arrests in which tasers were used. The y-axis denotes odds ratios of an officer shooting at a black civilian versus a white civilian. The x-axis denotes the period of years for which the odds ratios were calculated. We control for civilian gender, a quadratic in age, officer demographics, encounter characteristics, and missing indicators in all variables (i.e., all variables included in the final row of table 4). Year fixed effects are included in all regressions. Robust standard errors are reported in parentheses.

Are there racial differences in the timing of lethal force?—The above results, along with the results on use of force, are about racial differences on the extensive margin: whether or not an officer uses a particular type of force or decides to use lethal force on a suspect. Because of the richness of our OIS database, we can also investigate the intensive margin—whether there are racial differences in how quickly a police officer shoots a suspect in an interaction. In particular, given the narrative accounts, we create a dichotomous variable that is equal to one if a police officer reports that he shoots a suspect before being attacked and zero if he reports shooting the suspect after being attacked. These data are available for Houston as well as the other 15 locations where we collected OIS data. An important caveat to these data is that the sequence of events in a police-civilian interaction is subject to misreporting by police. Thus, the dependent variable is subjective.

Table 5 presents a series of estimates of racial differences in the timing of police shootings using the OIS data. The specifications estimated have the following form:

$$\ln\left(\frac{\Pr(\text{Shoot First}_{i,c,t})}{1 - \Pr(\text{Shoot First}_{i,c,t})}\right) = \text{Race}'_i\alpha + X'_{i,t}\beta + Z'_{c,t}\mathcal{T} + \nu_t + \psi_c + \varepsilon_{i,c,t},$$

where $\text{Shoot First}_{i,c,t}$ is a measure of whether a police officer reports shooting individual i , in city c , in year t , before being attacked. Standard errors, which appear below each estimate, are clustered at the location level unless otherwise specified.

The results from these specifications are consistent with our previous results on the extensive margin. Row a displays the results from the raw data. Blacks are 2.8 percent less likely to be shot first by police. Hispanics are slightly more likely. Neither coefficient is statistically significant. Adding suspect or officer demographics does not alter the results.³¹

Row d accounts for important context at the time of the shooting: for instance, whether the shooting happened during daytime or nighttime and whether the suspect drew a weapon or attacked the officer. Including these variables decreases the black coefficient to 0.232 (0.633), which is statistically significant. The Hispanic coefficient is similar in size but less precisely estimated. Adding whether the suspect was eventually found to have a weapon and its type or including location and year fixed effects only strengthens the results in the unexpected direction. Including all controls available, officers report that they are 46.8 percent less likely to discharge

³¹ We also estimate the “intensity” of force used in officer-involved shootings by estimating racial differences in the total number of bullets used in a given police shooting. The average number of bullets in officer-involved shootings involving blacks is 0.438 (0.805) more relative to shootings that involve nonblack, non-Hispanics. However, this coefficient is statistically insignificant (not shown in tabular form).

TABLE 5
RACIAL DIFFERENCES IN LETHAL USE OF FORCE, CONDITIONAL ON AN INTERACTION,
INTENSIVE MARGIN, OFFICER-INVOLVED SHOOTINGS

	Nonblack/Non-Hispanic Mean (1)	Black (2)	Hispanic (3)
a. No controls	.539	.972 (.114)	1.071 (.231)
b. + suspect demographics		.966 (.092)	1.042 (.247)
c. + officer demographics		.848* (.081)	.873 (.207)
d. + encounter characteristics		.692*** (.096)	.751 (.180)
e. + suspect weapon		.584*** (.071)	.621** (.143)
f. + fixed effects		.532*** (.049)	.536*** (.120)
Observations			1,399

NOTE.—This table reports odds ratios from logistic regressions. The sample consists of officer involved shootings from Dallas, Austin, nine Florida counties, Jacksonville, Houston, Los Angeles, Denver, and Seattle between 2000 and 2015. The dependent variable is based on who attacked first. It is coded as 1 if the officer attacked the suspect first and 0 if the suspect attacked the officer first. The omitted race is nonblacks and non-Hispanics. Col. 1 gives the unconditional average of contacts that resulted in an officer firing his gun. Col. 2 reports logistic estimates for black civilians. Each row corresponds to a different empirical specification. Row a includes solely racial dummies. Row b adds civilian gender and a quadratic in age. Row c adds controls for the split of races of officers present at the scene, whether any female officers were present, whether multiple officers were present, and the average tenure of officers at the scene. Row d adds controls for the reason the officers were responding at the scene, whether the encounter happened during daytime, and whether the civilian attacked or drew a weapon. Row e adds controls for the type of weapon the civilian was carrying. Row f adds city and year fixed effects. Each row includes missing indicators in all variables. Standard errors are clustered at the police department level and are reported in parentheses.

* Significant at the 1 percent level.
** Significant at the 5 percent level.
*** Significant at the 10 percent level.

their firearms before being attacked if the suspect is black. The Hispanic coefficient is strikingly similar (46.4 percent less likely).

Appendix table 8 explores the heterogeneity in the data across various subsamples: the racial composition of the responding unit, number of officers who arrive at a scene, whether or not officers report that the suspect clearly drew a weapon or “appeared” to draw a weapon, whether the officer was on duty, and the call type. The final panel provides results disaggregated by location.

Estimated race coefficients across call types—whether officers were dispatched because of a violent crime, robbery, auto crime, or other type of call—are all negative, if anything. This is particularly interesting in light of the potential selection into the sample of OIS cases discussed earlier. Indeed, the majority of police shootings in our data occur during violent

crimes or robberies, and on these call types, blacks are less likely to be shot at first, if anything.

One of the more interesting subsamples is whether or not a suspect “appeared” to have a weapon versus an officer indicating that it was clear he had a weapon. This dovetails with many of the anecdotal reports of police violence and is thought to be a key margin on which implicit bias, and the resulting discriminatory treatment, occur. Eberhardt et al. (2004) find that police officers detect degraded images of crime-related objects faster when they are shown black faces first.

Yet our data from the field seem to reject this lab-based hypothesis, at least regarding officer-involved shootings. The coefficient on black for the subsample who police report clearly drew their weapon first is -0.113 (0.024). The same coefficient estimated on the set of interactions in which police assumed an individual had a weapon is -0.023 (0.026). The Hispanic coefficients are nearly identical.

More generally, the coefficients are uncommonly consistent across all subsamples of the data. Of the six tests of equality performed in the table, only one is significant at the 1 percent level. We cannot detect racial differences in officer-involved shootings on any dimension.

VI. Interpretation

A number of stylized facts emerge from the analysis of the preceding sections. On nonlethal uses of force, there are racial differences—sometimes quite large—in police use of force, even after controlling for a large set of controls designed to account for important contextual and behavioral factors at the time of the police-civilian interaction. As the intensity of use of force increases from putting hands on a civilian to striking him with a baton, the overall probability of such an incident occurring decreases but the racial difference remains roughly constant. On the most extreme uses of force, however—officer-involved shootings with a taser or lethal weapon—there are no racial differences either in the raw data or when accounting for controls.

In this section, we explore the extent to which a model of police-civilian interaction that encompasses both information- and taste-based discrimination can successfully account for this set of facts. The model is an adaptation of Coate and Loury (1993a, 1993b).

A. *A Model of Police-Civilian Interactions*

Basic Building Blocks

Imagine a large number of police officers and a weakly larger population of civilians. Each police officer is randomly matched with civilians from

this population. Civilians belong to one of two identifiable groups, B or W . Denote by λ the fraction of W 's in the population. Police officers are assumed to be one of two types: "biased" or "unbiased." Let $\delta \in (0, 1)$ denote the fraction of biased police officers.

Nature moves first and assigns a cost of compliance to each civilian and a type to each police officer. Let $c \in [\underline{c}, \bar{c}]$ represent the cost to a civilian of investing in compliance. An alternative way to think about this assumption is that individuals contain inherent dangerousness, and those who are dangerous have higher costs of compliance.

After observing his cost, the civilian makes a dichotomous compliance decision, choosing to become either a compliant type or a noncompliant type, with no in between. Then, on the basis of this decision, nature distributes a signal $\theta \in [\underline{\theta}, \bar{\theta}]$ to police officers regarding whether or not a civilian is likely to comply.³² Next, the police officer observes θ and decides whether or not to use force, which we denote by $h \in \{0, 1\}$.³³

The distribution of θ depends, in the same way for each race, on whether or not a civilian has invested in compliance. This signal is meant to capture the important elements of initial interactions between police and civilians: clothing, demeanor, attitude, posture, and so on. Let $F_1(\theta)$ ($F_0(\theta)$) be the probability that the signal does not exceed θ , given that a civilian has invested in compliance (noncompliance), and let $f_1(\theta)$ and $f_0(\theta)$ be the related density functions. Define $\mu(\theta) \equiv f_0(\theta)/f_1(\theta)$ to be the likelihood ratio at θ . We assume that $\mu(\theta)$ is nonincreasing on $[0, 1]$, which implies that $F_1(\theta) \leq F_0(\theta)$ for all θ . Thus, higher values of observed θ are more likely if the civilian is compliant, and for a given prior, the posterior likelihood that a civilian will be compliant is larger if his signal takes a higher value.

Payoffs

For the civilian, payoffs depend on whether or not force is used on him and whether he chose to invest in compliance. Specifically, if force is used on the civilian, he receives a payoff of $-\gamma - c$ if he invested in compliance and $-\gamma$ if not. If force is not used on the civilian, he receives a payoff of $-c$ if he invests, and the payoff is normalized to zero if he did not invest.

³² This model is a simplified version of a more general model in which individuals invest in a "compliance identity" à la Akerlof and Kranton (2000) and then, in any given interaction with police, decide whether to comply or escalate. For those who have a compliance identity, there is an identity cost of escalation. This model is more intuitive but delivers the same basic results.

³³ We model the police officer's decision to use force rather than what type of force to use for analytical convenience and because for most of our analysis the dependent variable is whether or not to use force. Extending our analysis to allow for N potential uses of force does not alter the key predictions of the model.

It is assumed that police officers want to use force on civilians who are noncompliant and prefer not to use force on those who are compliant. In addition, we allow for “biased” police officers to gain utility from using force on B ’s.

Thus, for police officers, payoffs depend on their type, whether or not they use force, and whether or not the civilian is compliant. We begin with unbiased officers. If force is used, the officer’s payoff is $-K - \phi_F$ if the civilian is compliant and $\chi_F - \phi_F$ if the civilian is noncompliant. If no force is used, the officer receives a payoff of 0 if the civilian is compliant and $-\chi_{NF}$ if the civilian is noncompliant. These payoffs are identical for biased officers when they interact with W civilians.

When biased police officers interact with B civilians, they derive psychic pleasure from using force, independent of whether they are compliant or not. We represent this by τ , a positive term in the biased officer’s payoff when he uses force on B civilians. Note that this is similar to the taste parameter pioneered in Becker (1957).

Strategies

A civilian’s strategy is a mapping $I: [\underline{c}, \bar{c}] \rightarrow \{0, 1\}$. Without loss of generality, the civilian’s strategy can be represented by a cutoff point, c^* , such that the civilian will invest in compliance if and only if his cost is below c^* . A strategy for the police officer is a decision of whether or not to use force, conditional on what he can observe, $h: \{0, \tau\} \times [B, W] \times [\underline{\theta}, \bar{\theta}] \rightarrow \{0, 1\}$.

Expected Payoffs

Let $\pi \in [0, 1]$ denote the officer’s prior belief that a civilian will be compliant. Expected payoffs for the police officer are functions of her beliefs, her type, and the signal she receives. Given π and observed signal θ , she formulates a posterior probability (using Bayes’s rule) that the civilian will be compliant:

$$\Psi(\pi, \theta) \equiv \frac{\pi f_1(\theta)}{\pi f_1(\theta) + (1 - \pi)f_0(\theta)}.$$

The expected payoff of using force for an unbiased police officer (and, equivalently, a biased police officer when interacting with W ’s) is

$$\Psi(\pi, \theta)(-K - \phi_F) + [1 - \Psi(\pi, \theta)](\chi_F - \phi_F). \quad (2)$$

The expected payoff of using force for a biased officer interacting with B ’s is

$$\Psi(\pi, \theta)(-K - \phi_F) + [1 - \Psi(\pi, \theta)](\chi_F - \phi_F) + \tau. \quad (3)$$

Relatedly, the expected payoffs of not using force, for both types of officers, can be written as

$$-[1 - \Psi(\pi, \theta)](\chi_{NF}). \quad (4)$$

Combining equation (2) and equation (4), and using a bit of algebra, an unbiased officer uses force only if

$$\theta|\theta_{ub}^* \leq \min\{\theta|\Psi(\pi, \theta)(-K - \phi_F) + [1 - \Psi(\pi, \theta)](\chi_F + \chi_{NF} - \phi_F) > 0\}. \quad (5)$$

In words, equation (5) provides a threshold, θ_{ub}^* , such that for any θ below this threshold, unbiased officers always use force. Similarly, using the corresponding expected payoffs for a biased officer, one can derive θ_b^* .

Now, consider the civilian's expected payoff. Type W civilians receive $F_1(\theta_{ub}^*)(-\gamma) - c$ if they invest and $F_0(\theta_{ub}^*)(-\gamma)$ if they choose not to invest. When optimizing, a civilian will invest in compliance if and only if the cost of compliance is less than the net benefit of compliance. In symbols,

$$c \leq c_W^* \equiv \{F_0(\theta_{ub}^*) - F_1(\theta_{ub}^*)\}\gamma. \quad (6)$$

Similarly, B 's invest if

$$c \leq c_B^* \equiv \gamma\{\delta[F_0(\theta_{ub}^*) - F_1(\theta_{ub}^*)] + (1 - \delta)[F_0(\theta_b^*) - F_1(\theta_b^*)]\}. \quad (7)$$

Note—given that we assume $\delta > 0$ —it follows that $c_B^* < c_W^*$.

DEFINITION 1. An equilibrium consists of a pair (θ^*, π^*) such that each is a best response to the other.

B. Understanding the Data through the Lens of the Model

Assuming that the distribution of costs (c) and the signal (θ) are independent of race, racial disparities can be produced in this model in two (non-mutually exclusive) ways: different beliefs or different preferences.³⁴ To see this formally, suppose all racial differences were driven by information-based discrimination and there was no taste-based component. In this case, equation (3) simplifies to (2) and both B and W individuals' net benefit of investment becomes $[F_0(\theta_{ub}^*) - F_1(\theta_{ub}^*)]\gamma - c$. Thus, one needs differences in π to generate discriminatory equilibrium.

In contrast, one can also derive an equilibrium for cases in which we turn off the information-based channel and allow differences only through

³⁴ It is also plausible that racial differences arise because of differences in costs of compliance (for instance, through peer effects) or in the signal distributions. Incorporating these assumptions into the model is a trivial extension.

preferences. In this case, police officers observe investment decisions perfectly. When police officer bias is sufficiently large, any equilibrium will contain discrimination against B 's.

Distinguishing between these two cases, empirically, is difficult with the available data. In what follows, we attempt to understand whether the patterns in the data are best explained by an information-based or taste-based approach to discrimination—recognizing that both channels may be important.

Statistical Discrimination

To better understand whether statistical discrimination might explain some of the patterns in the data, we investigate two possibilities.³⁵ First, we explore whether racial differences in mean characteristics across police precincts predict racial differences in use of force. The key—untestable—assumption is that police officer beliefs about the compliance of a civilian— π in our model—are partly driven by local variation in variables such as education or income levels.³⁶

Appendix table 11 explores racial differences in any use of force—using the Stop and Frisk data—for various proxies for “dangerousness” including education, income, and unemployment. Education is represented by the fraction, by race, in each precinct of individuals with a high school diploma. Income is measured as median income. Unemployment is measured as the fraction of civilians in the labor force who are unemployed. For each of these variables, we take the difference between the white population and the black population and rank the precincts by this difference, individually. We then divide the data into terciles. The first tercile is always the one in which racial differences between our proxies are the lowest. The third tercile represents precincts in which there are relatively large racial differences on a given proxy.

Statistically larger racial differences in use of force for the third tercile (tercile 1 for unemployment), relative to tercile 1 or 2 (tercile 2 or 3 for unemployment), would be evidence consistent with statistical discrimination. This would imply that racial differences in use of force are correlated with racial differences in proxies for dangerousness. Appendix table 11 demonstrates no such pattern. The ordinary least squares coefficient of

³⁵ Appendix C considers the extent to which discrimination based on categories can explain the results (Fryer and Jackson 2008). We argue that categorical discrimination is inconsistent with the fact that black officers and white officers interact similarly with black civilians. See app. table 14.

³⁶ Ideally, one might use variables more directly correlated with dangerousness such as racial differences in crime rates, by precincts. Despite repeated formal Freedom of Information Act requests, the New York Police Department refused to supply these data.

having any force used on a black civilian versus a white civilian remains statistically the same across terciles.³⁷

A second prediction of the statistical discrimination model that is testable in our data is how racial differences in use of force change as signals about civilian compliance become more clear.³⁸ If statistical discrimination is the key driver of racial differences in use of force, the model predicts that as θ becomes perfectly predictive of compliance behavior, there will be no racial differences. We test this using officer-recorded data on the compliance behavior of civilians.

The NYC Stop and Frisk data contain officer-recorded information on the compliance of civilians during a stop. These variables include whether the civilians refused to comply with officers' directions, whether civilians verbally threatened an officer, whether they were evasive in their response to questioning, or whether they changed direction at the sight of an officer. If statistical discrimination is a key driver of racial differences, on the set of interactions in which officers report perfect compliance (and, to capture potentially important unobservables, the civilian was not arrested or was not guilty of carrying weapons or contraband), racial differences should be close to zero. And, on the set of interactions in which civilians engage in questionable behavior, racial differences should be statistically larger.

Figure 5 shows that even when we take perfectly compliant individuals and control for civilian, officer, encounter, and location variables, black civilians are 21.2 percent more likely to have any force used against them in an interaction compared to white civilians with the same reported compliance behavior. As the intensity of force increases, the odds ratio for perfectly compliant individuals decreases.

Ultimately, it is difficult to know if statistical discrimination is an important component of racial differences in use of force. Though our tests have quite limited power, we find no evidence that statistical discrimination plays an important role.

³⁷ We performed a similar exercise exploiting the variance across space in proxies for dangerousness (see app. tables 12A–12C for results). We also investigated whether more weight in the bottom quintiles of the distribution of our proxies predicted police use of force. These empirical exercises were meant as a partial test of Aigner and Cain (1977). We find no evidence of this sort of statistical discrimination on any of the dimensions tested.

³⁸ Another potential test of statistical discrimination was pioneered by Altonji and Pierret (2001). They investigate racial differences in wage trajectories, conditional on being hired. To the extent that statistical discrimination drives wage differences between racial groups, one would expect the wage trajectory for blacks to be higher than for whites—as employers learn. We performed a similar, though imperfect, test by estimating the probability that a civilian is arrested, conditional on force being used. Consistent with a discrimination story, on the lowest-level use of force, blacks and Hispanics are less likely to be arrested conditional on force being used. As the intensity of force increases, if anything, minorities are more likely to be arrested conditional on force being used.

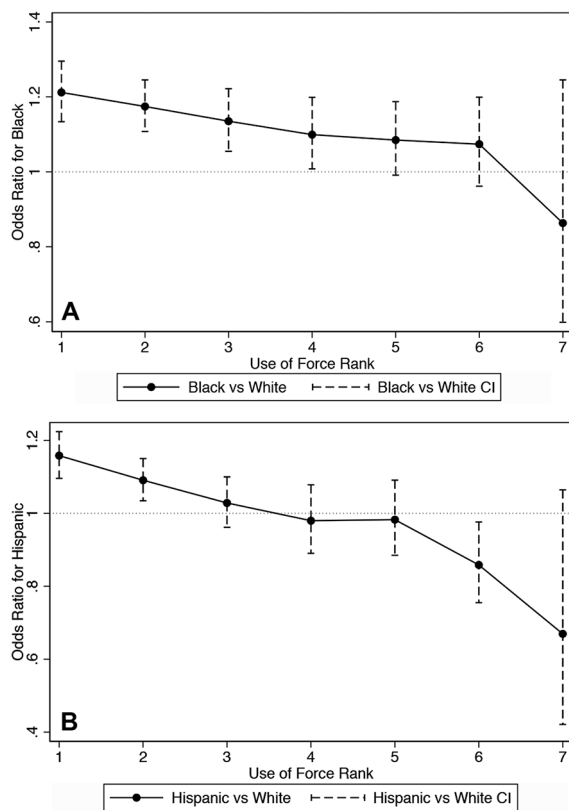


FIG. 5.—Odds ratios by use of force for perfectly compliant civilians (conditional on an interaction), NYC Stop, Question, and Frisk. These figures plot odds ratios with 95 percent confidence intervals from logistic regressions. For panel A, the y-axis denotes the odds ratio of reporting various uses of force for perfectly compliant black civilians versus perfect compliant white civilians. For panel B, the y-axis denotes the odds ratio of reporting various uses of force for perfectly compliant Hispanic civilians versus perfectly compliant white civilians. For both figures, the x-axis denotes different use of force types: 1 is an indicator for whether the police reported using at least hands or more severe force on a civilian in a stop and frisk interaction; 2 is for whether the police reported at least pushing a civilian to a wall or using more severe force; 3 is for whether the police reported at least using handcuffs or more severe force; 4 is for whether the police reported at least drawing a weapon on a civilian or using more severe force; 5 is for whether the police reported at least pushing a civilian to the ground or using more severe force; 6 is for whether the police reported at least pointing a weapon at a civilian or using more severe force; and, finally, 7 is for whether the police reported at least using pepper spray or a baton on a civilian. All force indicators are coded as 0 when the police report using no force in a stop and frisk interaction. The line plot is achieved by regressing the type of force on civilian race dummies, civilian gender, a quadratic in age, civilian behavior, whether the stop was indoors or outdoors, whether the stop took place during the daytime, whether the stop took place in a high-crime area or at a high-crime time, whether the officer was in uniform, civilian ID type, whether others were stopped during the interaction, and missing indicators in all variables. Precinct and year fixed effects were included in all regressions. Standard errors are clustered at the precinct level.

Taste-Based Models of Discrimination

Similar to any large organization, police departments surely have individuals who hold biased views toward minority citizens, and those views may manifest themselves in biased treatment of individuals based solely on their race. Yet, as Becker (1957) argued, individual discrimination does not necessarily equate to market (or systemic) discrimination.

Taste-based discrimination is consistent with the data from the direct regression approach on nonlethal uses of force if, among those who discriminate, the preference for discrimination is greater than the expected costs of wrongly using force. In other words, the expected price of discrimination is not large enough—through either low penalties or low probabilities of detection—to alter behavior of those who have biased preferences. This model is also consistent with the lack of racial differences in officer-involved shootings if there is a discrete increase in the costs of being deemed a discriminator, relative to the costs incurred with nonlethal uses of force.³⁹

Below, we explore the extent to which two additional implications of the taste-based channel of our model are borne out in the data. The first uses the predictions on average versus marginal returns of compliant behavior. The second is inspired by the seminal work in Knowles et al. (2001) and Anwar and Fang (2006).

In any equilibrium model of discrimination, officer behavior influences the incentive to invest in compliance behavior. This is made explicit in equations (6) and (7). Figure 5 provides some suggestive evidence that the returns to compliance may be different across races. We can test this a bit more directly. One issue in this setting, which does not arise in labor markets, is that it is not obvious how to aggregate noncompliance into a monotonic index. From a police officer's perspective, it may be considered more dangerous if a civilian shouted verbal threats than if he refused to comply with an officer's directions or if he was evasive during questioning. A simple aggregation of the number of noncompliant activities is likely misleading.

To sidestep this important potential issue of aggregating noncompliance, we create an index equal to 1 if a civilian changes direction at the sight of an officer, 2 if a civilian is noncompliant on any other, but not all, dimensions of measured compliance, and 3 if a civilian is noncompli-

³⁹ While purely anecdotal, in police departments across the country, any officer-involved shooting—no matter how “justified”—results in the temporary confiscation of the officer's weapon until an investigation of the incident is complete. This is a potentially high cost relative to other nonlethal uses of force. Moreover, in informal interviews with dozens of police officers in Boston, Cambridge, Camden, and Houston, almost all police officers described pulling the trigger of their weapon as a “life-altering event.”

ant on all four dimensions we can measure. The regression estimated, then, is whether or not an officer uses any force—accounting for our full set of controls and including our measure of noncompliance interacted with race. Racial differences in the marginal return to noncompliance behavior would manifest themselves in statistically different coefficients on the compliance variable. For a given race, adding both the race coefficient and the interaction term with compliance behavior provides an estimate of the net benefit of investment (eqq. [6] and [7]).

The results of this exercise (not shown in tabular form) are consistent with racial differences in police use of force being driven by taste-based discrimination. Black civilians have marginal returns to compliance statistically similar to those of white civilians. In other words, the probability of force being used as θ increases is statistically identical between blacks and whites. Yet black civilians always have a higher likelihood of force being used on them compared to white civilians, for all θ . Further, the net benefit of investment in compliance is lower for blacks relative to whites. This is precisely what the model predicts if racial animus is an important factor in explaining racial differences in use of force.

We conclude our statistical analysis by developing a test for discrimination based on those by Knowles et al. (2001) and Anwar and Fang (2006) to complement the direct regression approach described in the previous sections. Knowles et al. test for racist preferences by looking at officers' success rate of searches across races. Their model assumes that police maximize the number of successful searches net of the cost of searching motorists. If racial prejudice exists, then the cost of searching drivers will be different across races. This, in turn, implies that the rate of successful searches will be different across races.

Anwar and Fang (2006) build on the theory of Knowles et al. (2001), arguing that the Knowles et al.'s results might not hold if police officers are nonmonolithic in their behavior. They test this by investigating search rates of civilians of a particular race, across officer races. Under the null hypothesis that none of the racial groups of officers have relative racial prejudice, it must be true that the ranking of search rates for white civilians across officer races is the same as the ranking of search rates for black civilians across officer races.

We adopt this approach by investigating whether or not a suspect was eventually found to have a weapon during the interaction with police. In other words, we calculate the probability, for each race, that a suspect has a weapon conditional on being involved in an officer-involved shooting. Given the level of detail in our data, one can perform this test for weapons generally—guns, knives or other cutting objects, or assault weapons—or for guns, including pistols, rifles, or semiautomatic machine guns, specifically. Moreover, following the insights in Anwar and Fang (2006), we disaggregate the data by officer race.

The null hypothesis is no racial discrimination in officer-involved shootings. The null could be rejected in several ways. First, according to Knowles et al. (2001), the null could be rejected if the fraction of suspects carrying weapons or firearms is different across suspect races. Second, according to Anwar and Fang (2006), the null could be rejected if the ranking of “being armed” rates for black suspects across officer races is different from the ranking of being armed rates for white suspects.

Consistent with our direct regression approach and the findings in Knowles et al. (2001) and Anwar and Fang (2006), we fail to reject the null of no discrimination. The data are displayed in table 6. For white officers, the probability that a white suspect who is involved in an officer-involved shooting has a weapon is 84.1 percent. The equivalent probability for blacks is 80.7 percent: a difference of almost 4 percent, which is not statistically significant. For black officers, the probability that a white suspect who is involved in an officer-involved shooting has a weapon is surprisingly lower, 57.1 percent. The equivalent probability for black suspects is 73.0 percent. The only statistically significant differences by race demonstrate that black officers are more likely to shoot unarmed whites relative to white officers.

We perform a similar exercise for nonlethal uses of force, recognizing that as the use of force gets less extreme, the application of that force and whether or not a suspect has a weapon is more tenuous. For instance, investigating racial differences in whether or not officers use hands on civilians who are unarmed is not a valid test of discrimination as there are myriad legitimate reasons for police officers to place hands on civilians who are unarmed. Yet racial differences in the use of a baton—after accounting for suspect behavior—seem less justifiable. Unfortunately, where to draw the line on the continuum of potential uses of force is ad hoc. Thus, we present

TABLE 6
FRACTION WEAPON FOUND, CONDITIONAL ON BEING
IN AN OFFICER-INVOLVED SHOOTING

	Civilian White (1)	Civilian Black (2)	<i>p</i> -Value (3)
Officer white	.841 (.028)	.807 (.026)	.376
Officer black	.571 (.137)	.730 (.056)	.246
<i>p</i> -value	.011	.186	

NOTE.—This table presents results for the Anwar and Fang (2006) test. Col. 1 presents the fraction of white civilians carrying weapons in the OIS data set. Col. 2 presents the fraction of black civilians carrying weapons in the OIS data set. Col. 3 displays the *p*-value for equality of means in cols. 1 and 2. The first row presents the fractions in which the majority of officers present during the encounter were white. The second row presents the fractions in which the majority of officers present during the encounter were black.

our modified Knowles et al. test for all uses of force while acknowledging that for the low-level uses, it does not seem appropriate.

Appendix table 13 presents these results. Each row is a different level of force that begins with “at least hands” and increases in severity of force until “use of pepper spray or baton.” Column 1 contains the white mean. Columns 2 and 3 display the coefficient on black and Hispanic, respectively. Column 4 displays the number of observations, which range from over 1 million for the use of hands to 1,745 for the use of pepper spray or baton.

Blacks are 1.0 (0.1) percentage points less likely to have a weapon, conditional on a police officer using any force. Hispanics are 0.6 (0.1) less likely to have a weapon. Both are statistically significant. Interestingly, on the two most severe nonlethal uses of force, the probability that a weapon is found—conditional on force being used—is statistically identical across races. Taken at face value, these data are consistent with discrimination against minorities on the lowest-level uses of nonlethal force.

VII. Conclusion

The issue of police violence and its racial incidence has become one of the most divisive topics in American discourse. Emotions run the gamut from outrage to indifference. Yet very little data exist to understand whether racial disparities in police use of force exist or might be explained by situational factors inherent in the complexity of police-civilian interactions. Beyond the lack of data, the analysis of police behavior is fraught with difficulty including, but not limited to, the reliability of the data that do exist and the fact that one cannot randomly assign race.

With these caveats in mind, this paper takes first steps into the treacherous terrain of understanding the nature and extent of racial differences in police use of force and the probability of police interaction. On nonlethal uses of force, there are racial differences—sometimes quite large—in police use of force, even after accounting for a large set of controls designed to account for important contextual and behavioral factors at the time of the police-civilian interaction. Interestingly, as use of force increases from putting hands on civilians to striking them with a baton, the overall probability of such an incident occurring decreases dramatically, but the racial difference remains roughly constant. Even when officers report that civilians have been compliant and no arrest was made, blacks are 21.2 percent more likely to endure some form of force in an interaction. Yet, on the most extreme use of force—officer-involved shootings—we are unable to detect any racial differences either in the raw data or when accounting for controls.

We argue that these facts are most consistent with a model of taste-based discrimination in which police officers face discretely higher costs for officer-involved shootings relative to nonlethal uses of force. This model

is consistent with racial differences in the average returns to compliant behaviors, the results of our tests of discrimination based on Knowles et al. (2001) and Anwar and Fang (2006), and the fact that the odds ratio is large and significant across all intensities of force—even after accounting for a rich set of controls. In the end, however, without randomly assigning race, we have no definitive proof of discrimination. Our results are also consistent with mismeasured contextual factors.

As police departments across America consider models of community policing such as the Boston Ten Point Coalition, body-worn cameras, or training designed to purge officers of implicit bias, our results point to another simple policy experiment: increase the expected price of excessive force on lower-level uses of force. To date, very few police departments across the country either collect data on lower-level uses of force or explicitly punish officers for misuse of these tactics.

The appealing feature of this type of policy experiment is that it does not require officers to change their behavior in extremely high-stakes environments. Many arguments about police reform fall victim to the “my life versus theirs, us versus them” mantra. Holding officers accountable for the misuse of hands or pushing individuals to the ground is not likely a life or death situation and, as such, may be more amenable to policy change.

The importance of our results for racial inequality in America is unclear. It is plausible that racial differences in lower-level uses of force are simply a distraction and movements such as Black Lives Matter should seek solutions within their own communities rather than changing the behaviors of police and other external forces.

Much more troubling, because of their frequency and potential impact on minority belief formation, is the possibility that racial differences in police use of nonlethal force have spillovers on myriad dimensions of racial inequality. If, for instance, blacks use their lived experience with police as evidence that the world is discriminatory, then it is easy to understand why black youths invest less in human capital or black adults are more likely to believe that discrimination is an important determinant of economic outcomes. Black Dignity Matters.

Appendix

Table 1 Calculations

To obtain the number reported in panel A, column 1, of table 1, we go through the following steps: For each precinct, calculate the fraction of stops that are black and the corresponding fraction for whites; for each precinct, calculate the fraction of males aged 18–34 in the population who are black and the corresponding fraction for whites; regress the fraction of stops that are black on the fraction of males aged 18–34 who are black (with no constant) for all 77 pre-

cincts. The beta coefficient on the dependent variable shows the representation of “at-risk” blacks in stops. Conduct the same regression for whites and store that beta coefficient as the representation of at-risk whites in stops; finally, divide the beta coefficient for blacks by the beta coefficient for whites. To obtain the numbers in panel A, columns 2 and 3, we go through the following steps: For each year, calculate the fraction of stops that are black and the corresponding fraction for whites for New York City; for each year, calculate the fraction of arrestees for the 10 most egregious felonies and misdemeanors for column 2 and the fraction of arrestees for the six most egregious felonies only for column 3 for blacks and whites for New York City; regress the fraction of stops that are black on the fraction of arrestees that are black (with no constant) for all 6 years of data and store the beta coefficient and do the same regression for whites and store the beta coefficient; finally, divide the beta coefficient for blacks by the beta coefficient for whites. Columns 4 and 5 of panel A report odds ratios on blacks from logistic regressions of involuntary or voluntary interaction on civilian race, other civilian demographics, and year. To obtain the numbers in panel A, columns 6–8, we go through the following steps: Calculate the fraction of blacks in officer-involved shootings and divide it by the fraction of blacks in the male population aged 18–34 in Houston/fraction of blacks in all felony and misdemeanor arrests/fraction of blacks in all felony arrests; calculate the same fraction for nonblacks; and finally divide the black fraction by the nonblack fraction.

References

- Aigner, D. J., and G. G. Cain. 1977. “Statistical Theories of Discrimination in Labor Markets.” *Indus. and Labor Relations Rev.* 30 (2): 175–87.
- Akerlof, G. A., and R. E. Kranton. 2000. “Economics and Identity.” *Q.J.E.* 115 (3): 715–53.
- Altonji, J. G., and C. R. Pierret. 2001. “Employer Learning and Statistical Discrimination.” *Q.J.E.* 116 (1): 313–50.
- Anwar, S., and H. Fang. 2006. “An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence.” *A.E.R.* 96 (1): 127–51.
- Arthur, Rob, Taylor Dolven, Keegan Hamilton, Allison McCann, and Carter Sherman. 2017. “Shot by Cops and Forgotten.” VICE News, December 11. news.vice.com/en_us/article/xwv3a/shot-by-cops.
- Becker Gary S. 1957. *The Economics of Discrimination*. 2nd ed. Chicago: Univ. Chicago Press.
- . 1993. “Nobel Lecture: The Economic Way of Looking at Behavior.” *J.P.E.* 101 (3): 385–409.
- Coate, S., and G. C. Loury. 1993a. “Antidiscrimination Enforcement and the Problem of Patronization.” *A.E.R.* 83 (2): 92–98.
- . 1993b. “Will Affirmative-Action Policies Eliminate Negative Stereotypes?” *A.E.R.* 83:1220–40.
- Cose, E. 1993. *The Rage of a Privileged Class: Why Are Middle-Class Blacks Angry? Why Should America Care?* New York: HarperPerennial.
- Coviello, D., and N. Persico. 2015. “An Economic Analysis of Black-White Disparities in NYPD’s Stop and Frisk Program.” *J. Legal Studies* 44 (2): 315–60.
- Eberhardt, J. L., P. A. Goff, V. J. Purdie, and P. G. Davies. 2004. “Seeing Black: Race, Crime, and Visual Processing.” *J. Personality and Soc. Psychology* 87 (6): 876–93.

- Fridell, L. A. 2004. "By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops." Police Executive Research Forum, Washington, DC.
- Fryer, Roland G., Jr. 2018. "Reconciling Results on Racial Differences in Police Shootings." *A.E.R. Papers and Proc.* 108:228–33.
- Fryer, R., and M. O. Jackson. 2008. "A Categorical Model of Cognition and Biased Decision Making." *B. E. J. Theoretical Econ.* 8 (1).
- Gelman, A., J. Fagan, and A. Kiss. 2012. "An Analysis of the New York City Police Department's 'Stop-and-Frisk' Policy in the Context of Claims of Racial Bias." *J. American Statis. Assoc.* 107 (479): 813–23.
- Kindy, Kimberly, Marc Fisher, Julie Tate, and Jennifer Jenkins. 2015. "A Year of Reckoning: Police Fatally Shoot Nearly 1,000." *Washington Post*, December 26. <https://www.washingtonpost.com/sf/investigative/2015/12/26/a-year-of-reckoning-police-fatally-shoot-nearly-1000/>.
- Knowles, J., N. Persico, and P. Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *J.P.E.* 109 (1): 203–29.
- Miller, J., R. C. Davis, N. J. Henderson, J. Markovic, and C. W. Ortiz. 2004. "Public Opinions of the Police: The Influence of Friends, Family and News Media." Vera Inst. Justice, New York.
- Mincer, J. 1958. "Investment in Human Capital and Personal Income Distribution." *J.P.E.* 66:281–302.
- Neal, D. A., and W. R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *J.P.E.* 104:869–95.
- O'Neill, J. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." *J. Econ. Perspectives* 4:25–45.
- Ridgeway, G. 2007. "Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices." Technical report, RAND Corp., Santa Monica, CA.
- Ridgeway, G., and J. M. MacDonald. 2009. "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops." *J. American Statis. Assoc.* 104 (486): 661–68.
- Ross, Cody. 2015. "A Multi-level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014." *PLoS ONE* 10 (11).
- Schneider, A. 1977. "Portland (OR) Forward Record Check of Crime-Victims Final Report, December 1977." Oregon Res. Inst., USA, Eugene.
- Swaine, Jon, Oliver Laughland, Jamiles Larteyand, and Ciara McCarthy. 2015. "Young Black Men Killed by US Police at Highest Rate in Year of 1,134 Deaths." *Guardian*, December 31. <https://www.theguardian.com/us-news/2015/dec/31/the-counted-police-killings-2015-young-black-men>.