

# Everyday Risk Disparate Exposure and Racial Inequality in Police Violence

Laurel Eckhouse<sup>†</sup>

April 15, 2018

## Abstract

Scholarship on police violence has focused on identifying cognitive and psychological causes of implicit bias among police officers. This research strategy poses both analytic and policy problems. The theory of disparate exposure provides an alternative explanation for racial inequalities in police violence. I develop this theory, then test it using an interrupted time series from New York City. A sudden procedural change reduced the exposure of black civilians to casual stops without changing the personnel, implicit bias, or other qualities of the NYPD. Exposure reductions eliminated around 40 incidents of using force against African Americans per day in a 100-day bandwidth; estimates on the full data set show a reduction of 185 uses of force each day. National data suggests that these findings hold beyond New York. This study is the first to examine the effect of disparate exposure on racial inequalities in police violence.

---

\*I am grateful to the National Science Foundation Graduate Research Fellowship Program, the Horowitz Foundation for Social Policy, the Human Rights Data Analysis Group, and the Travers Department of Political Science at UC Berkeley for support for this research. For helpful feedback and assistance, I also thank Miguel de Figueiredo, Joy Milligan, Gabe Lenz, Amy Lerman, Ian Dickson, Renee Perry, Jennifer Cryer, Anton Strezhnev, Ryan McMahon, Jonathan Mummolo, Fridolin Linder, and Alex Coppock.

<sup>†</sup>Assistant Professor at the University of Denver. laurel.eckhouse@du.edu

On July 6, 2016, Philando Castile was pulled over in Falcon Heights, Minnesota. His girlfriend and her four-year-old daughter were with him in the car. Within minutes of the beginning of the stop, police officer Jeronimo Yanez had shot Castile four times (LaFraniere and Smith, 2016). Nationwide protests erupted in the wake of Castile's death – and the previous day's shooting of Alton Sterling in Louisiana. Dozens of protesters were arrested in the Twin Cities, and at protests around the country. As reporters uncovered the details, they learned that it was at least the 49th time Castile had been pulled over in 13 years (LaFraniere and Smith, 2016).

Journalists reported on Castile's many stops as an indicator of the heavy weight of law enforcement borne by low-income black communities. Castile's encounters with law enforcement, almost entirely over minor issues such as broken taillights, cost him over six thousand dollars in fines (Peralta and Corley, 2016). An extensive literature documents the consequences of these routine stops for African Americans: fines and fees paid, but also time spent in court and waiting rooms, suspended licenses, warrants, civic disengagement, lost jobs, and evictions (Goffman, 2014; Weaver and Lerman, 2010; Brayne, 2014; Lerman and Weaver, 2014a; Harris, 2016; White, 2015).

Each encounter, though, also carries the risk of escalation. Castile was stopped for a broken taillight; the officer also thought he resembled a robbery suspect. Castile was carrying a firearm – one he was licensed to carry under Minnesota state law, and one which he informed the officer of immediately (LaFraniere and Smith, 2016). Despite the law-abiding character of his behavior and the minor causes of the stop, Castile lost his life.

Over the last several years, scholars and journalists hastened to develop adequate measures of police violence, and to understand the nature and origins of the disproportionate risk of police violence faced by black and Native Americans, especially young men. The findings were stark. Federal records drastically understate the number of police killings (Ball, 2016). As many as one in five homicides committed by strangers are committed by police (Ball, 2016) and the death rates for black and Native Americans are over twice the death rates for white Americans. Unarmed black Americans are nearly five times as likely to be shot by police as unarmed white Americans (Zimring, 2017). In his introduction to the Department of Justice investigation of the Ferguson

police department, Shaw points out: “Police killings of unarmed individuals are, unfortunately, not uncommon. While the facts of each case are different, there is a numbing familiarity when an unarmed black boy, teenager, or man is killed by a police officer. A well-worn script unfolds after each death”, with the result that police killings rarely end in charges, convictions, or effective federal intervention (Shaw et al., 2015).

Yet when scholars, journalists, and activists tried to unpack the causes of this racial disproportion in police violence, they focused largely on the moment of decision. Police, like other Americans, are more likely to see “gun” in an ambiguous object if the person carrying it is black (Greenwald, Oakes and Hoffman, 2003). Some studies find that police, like other Americans, are more likely to shoot in experimental situations when the target is black (Greenwald, Oakes and Hoffman, 2003; Plant and Peruche, 2005; Correll et al., 2007). Other studies complicate the findings, suggesting that police may actually be slower to shoot unarmed black suspects than unarmed white suspects, despite findings of implicit bias and increased threat assessment among the officers in the experimental population (James, James and Vila, 2016), though these findings may be due to an observer effect or other mismatches between the experimental and real-world decision environment (Roussell et al., 2017). These cognitive biases are a substantial, if unsettled, concern; however, they are also difficult to change (Sim, Correll and Sadler, 2013; Eberhardt et al., 2004). Moreover, as I show in this paper, equalizing the probability of being shot in any given interaction would still leave tremendous racial inequality.

In focusing on how police react in highly charged situations where they are considering using force, scholars separate police homicide from the rest of the carceral state. In fact, these extreme instances of state violence have everything to do with the normal operation of the criminal justice system, and with the most mundane interactions between citizens and the criminal justice system.

Police shootings are the most high-profile type of police force, but the scale of non-fatal force dwarfs them in impact. The best current estimates suggest that over 1000 people are killed each year by police, but the Bureau of Justice Statistics estimates that 715,500 Americans each year have encounters where force is threatened or used. Non-lethal force can be tremendously politically

consequential (consider, for example, the riots in Los Angeles that followed the 1992 acquittal of police in the beating of Rodney King). In this paper, I use “police violence” to mean any use of physical force by a police officer.

Given a fixed risk of violence in any encounter with police – even a risk that is fixed at the individual level, rather than at the population level – an individual person’s lifetime risk of experiencing violence is a function both of the probability of being shot in any given encounter and the number of total encounters. Thus, racial disparities in the probability of experiencing violence by police can originate not only in the officer’s decision to use force, but in any of the prior steps that led to the interaction. Focusing on *cognitive bias* among officers misses the key role of *structural bias* in policing: the “scripts of suspicion” and institutionally biased practices that lead to substantial racial inequalities in policing *even before* cognitive bias has the opportunity to play a role within interactions.

In this paper, I develop a new theory of the causes of racial inequality in police use of force: the theory of *disparate exposure*. The paper’s main theoretical contribution is to identify disparate exposure – racial inequalities in the probability of police contact – as the central cause of racial inequality in police use of force. This theory connects incidents of police force to the broader carceral state. I test this theory by examining a discontinuous change in policing by the New York Police Department, which reduced disparate exposure without changing cognitive biases. When structural inequalities in the number of stops experienced by black New Yorkers changed, black New Yorkers’ risk of experiencing police violence fell dramatically. By focusing on a discontinuous change covered by highly granular administrative data, this paper can provide causal identification under plausible assumptions. During the 100-day bandwidth after the discontinuous change, the reduction in stops of black civilians prevented on the order of 40 uses of force *each day*.

# 1 The Origins of Inequality

For an unarmed black civilian, the risk of being shot by police is between 3.49 and 5 times the risk for an unarmed white civilian (Ross, 2015; Zimring, 2017). In many US counties, the risk for black civilians approaches 20 times the risk of white civilians (Ross, 2015). The death rates for African Americans are 2.3 times the death rates for whites. Native Americans are killed at 2.06 times the rate at which whites are killed (Zimring, 2017).

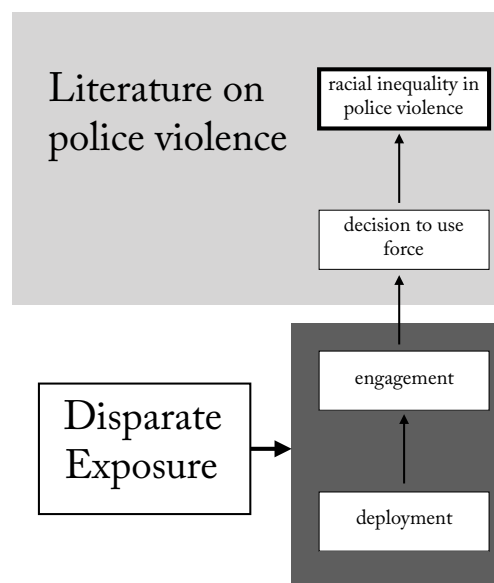


Figure 1: The Logic of Disparate Exposure

What causes racial inequality in police violence? One possible cause – the central one in the emerging scholarship on police violence – is that police make biased decisions about when to use a gun, a taser, or simple physical force. Yet, by the time the officer is choosing whether to use force on a particular individual, the officer has gone through many steps – training, deployment, and the decision to stop – that can produce racial disproportion *with no cognitive bias in the decision to use force*. As I show later, this disparate exposure is actually more significant than within-encounter

disproportions as an explanation for racial inequality in police use of force.

Figure 1 illustrates this process. At three key stages in policing – deployment, engagement, and the decision to use force – opportunities for disproportion enter and compound. Deployment and engagement together produce what I call *disparate exposure*. Focusing on the individual officer's decision to use force or not diverts attention from structural decisions over which individual officers exert little control, and from the racialized structures of policing, to individual cognition. By turning attention instead to *stops*, I focus on the racialized outcomes that result from institutional practices (Epp, Maynard-Moody and Haider-Markel, 2014). As I show later in the paper, cognitive bias in the force decision appears to account for far less of the racial disproportion in police use of force – lethal and not – than the preceding elements that lead to exposure disparities.

First, officers are deployed to a particular location. Ample evidence suggests that police are disproportionately deployed to low-income black and brown neighborhoods (Goel, Rao and Shroff, 2016; Moskos, 2009; Fagan et al., 2009). Drug enforcement and discretionary stops, in particular, are concentrated in what Soss and Weaver call “race-class subjugated communities” (Soss and Weaver, 2017; Lum and Isaac, 2016). Over the last four decades, “public authorities poured their expanded policing resources into a suite of new techniques such as zero tolerance policing, ‘command and control’ operations, order maintenance, ‘hot spots’ policing, saturation policing, and interventions based on the SARA (Scanning, Analysis, Response, and Assessment) model, all of which gave rise to a higher volume of lower quality arrests and convictions. Thus, as new policing models proliferated in the 1980s, 1990s, and 2000s – trumpeted through a string of tactical campaigns titled ‘Operation’ such and such – the core preoccupation of policing consolidated around the elimination of disorder and the regulatory enforcement of codes against disordered people and places” (Soss and Weaver, 2017). In the 1990s, the nationwide spread of data-driven policing (particularly Compstat) led to a further increase in minor arrests. Like other, similar technocratic interventions, Compstat was justified by the Broken Windows theory of policing, which held that policing disorder would reduce serious violent crime. More recent evidence, though, suggests that crime arises from private conflict, not public disorder (O’Brien and Sampson, 2015). Moreover,

Compstat did not lead to a significant decline in crime.

Today, intensive policing is the norm in race-class subjugated communities (Goffman, 2014; Soss and Weaver, 2017) but rare in wealthier neighborhoods (Moskos, 2009; Goel, Rao and Shroff, 2016). These deployment patterns mean that police officers are more likely to be working near black and brown Americans than near whites. Police training, management, and promotion expectations teach officers to see stops and arrests as indicators of productivity (Goel, Rao and Shroff, 2016). Proximity to race-class subjugated communities means not only that police may be more likely to observe any illegal behavior, but that as they seek out occasions to show productive work according to their organizational benchmarks, they will disproportionately engage with and arrest people of color – without individual cognitive bias at play. Deployment itself can lead to disparate exposure with no cognitive bias.

Second, police officers must choose whether to engage with a particular individual, with or without bias. Assessing bias in the decision to stop is complex, but the available evidence suggests that, even accounting for neighborhood deployment patterns, police are more likely to stop people of color and especially African Americans (Goel, Rao and Shroff, 2016). In a study of Kansas City area traffic stops, black drivers were 270% more likely to be stopped for “investigatory” purposes. These stops are often not the result of overt racial animosity. Rather, they are planned opportunities to investigate, understood by many police officers as “among their most effective tools for finding and arresting criminals and preventing crimes.” Since “even people who are opposed to racism often implicitly perceive others in racially stereotypical ways.... implicit negative racial stereotypes help to support punitive practices like the investigatory stop. In turn, these practices contribute to racial disparities in who is stopped[.]” (Epp, Maynard-Moody and Haider-Markel, 2014). Whether they originate in institutional recommendations, implicit bias, or overt animosity, racial differences in stops also contribute to disparate exposure.

Only after these institutional and individual decisions have led an officer to engage with a specific individual does the decision to use force (or not) occur. If exposure to police interaction occurs at disproportionate levels for African Americans, cognitively unbiased decisions to use force *in the*

*moment* would nevertheless result in more African Americans experience police violence.

While black Americans are 2.3 times as likely to be killed by police as whites generally (Zimring, 2017), after accounting for disparate exposure (using arrest data, which likely underestimates the extent of disparities in exposure) they are only 1.6 times as likely to be killed. Setting the number of police killings per arrest – the measure of cognitive bias in shootings – equal for blacks and whites would have meant 101 fewer deaths of black Americans in 2015. Fixing the black arrest rate at the white arrest rates, with no change to the race-specific ratio between killings and arrests, would eliminate 167 deaths – 67% more. Disparate exposure plays a very important role in predicting racial disparities in police shootings.<sup>1</sup>

For police reformers, this is encouraging news. Changing individual-level officer attitudes and behaviors, especially subconscious ones, is a difficult task. We have little evidence that training programs effectively diminish implicit bias, especially given the socializing effects of police work environments (Moskos, 2009; Christie, Petrie and Timmins, 1996; Paluck and Green, 2009). Even diversifying police forces – a major push over the last several decades, and one supported by black mayors (Saltzstein, 1989) – has not solved racial disparities in policing.

Changing deployment is a different matter. Historically, scholars have expressed concern that police activities are difficult to monitor. Police are street-level bureaucrats at the end of a loosely coupled chain of policy makers and implementers, and must engage independently with situations that remain unobserved by voters, police chiefs, and even their direct supervisors (Wilson, 1978). When policies direct police to change their observed work products, though, the evidence suggests that police often do so. They make more or fewer stops (Moskos, 2009; Mummolo, 2017) and change the composition of those stops to reflect departmental priorities. Cities have many policy levers available to address the ways their deployment and interaction policies produce disparate exposure. Even if implicit bias drives stop decisions that contribute to disparate exposure, we have reason to think that this may be a more fruitful point of intervention than the fast-paced, emotionally charged context of police stops.

---

<sup>1</sup>These calculations use data on police killings from *The Washington Post* and arrest data from the FBI Uniform Crime Reports.



In the sections that follow, I discuss the challenges of effectively measuring both police violence and exposure. I identify a data set on New York City’s Stop, Question, and Frisk (*Terry* stops) program which helps overcome these data problems. This data suggests that when changes in police practice reduce exposure, racial disparities in the use of force also decline. A discontinuous, policy-driven change in the frequency of SQF/*Terry* stops massively reduced the exposure of black civilians in New York City to police use of force – even though the probability of force being used in any individual stop actually rose. These findings suggest that scholars, activists, and policymakers concerned about police violence should look to the disparate exposure of black civilians to police interactions as a key factor in racial inequalities in police violence.

## **1.1 Disentangling Exposure and Bias: the Problem of Reference Populations**

Fryer’s 2016 working paper on police use of force attempts to place police gun violence in the context of other police actions. In order to determine whether police are more or less likely to shoot black suspects, he compares shootings in Houston to a reference set of other types of interactions. This raises a key issue: what is the appropriate reference population? Fryer argues for using “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 arrest.” (Fryer Jr, 2016)

However, many recent high profile shootings do not arise from this reference population. Philando Castile in particular – who was shot minutes into a traffic stop for a broken taillight – would not have been charged with any of these based on the observed interactions. The same is true for many other high-profile police shootings. Moreover, the final three (‘resisting arrest’, ‘evading arrest’, and ‘interfering in arrest’) are highly fungible categories to which people can easily be assigned based on the outcome of the interaction. Police can list someone as having ‘resisted arrest’ to justify a shooting. For this reason, researchers are likely to observe many cases of ‘resisting’ or ‘evading’ arrest among individuals who are shot, but these designations can be selected *after* the

shooting does or does not take place. That is, the civilian may resist arrest and thus be shot; or the causal arrow may run the other direction, so that the civilian is shot and, as a result, recorded as resisting arrest.

By sampling control cases from the same codes used to describe *justified* lethal force, Fryer accepts the narrow reference population designated by police, and omits control versions of *unjustified* cases as well as control versions of cases that are justified using post-hoc rationalizations. This creates serious problems in assessing whether bias actually exists. Fryer is comparing the rare positives in the sparse data to a sample, not of all cases which could lead to shootings, but to cases which are particularly likely to appear justified.

Considering all police contacts as the reference population provides one solution. Every contact carries with it the possibility of escalation, and numerous police killings began as traffic stops or interactions over non-serious offenses. Even so, if people of color are more likely to experience police contact for the same behavior, as national survey data on drug use suggests, the population *arrests of black Americans* may in fact be a less ‘dangerous’ population than *arrests of white Americans*; the same is true for any individual category of arrests. This type of bias in the composition of the population would make the estimates here a lower bound on the influence of individual, internal racial bias on racial bias in shootings.

An ideal data set for disentangling disproportion and bias would begin with the complete universe of police-civilian interactions with all initial intentions. Ideally, both the initial purpose of the stop and the outcome would be recorded: arrest or no arrest, all types of use of force, etc. Even more ideally, the data set would be sufficiently granular to identify the effects of policy changes on police strategies. No such data set exists at the national level. Indeed, there is not even a complete dataset of stops; the national analysis above showing the role of arrests in disparate exposure misses non-arrest stops and interactions. However, the New York Police Department has collected data that provides important leverage on the interaction between disparate exposure and bias within the NYPD.

## 2 New York: Policy Change and Disparate Exposure

On March 5, 2013, the New York City Police Department undertook a reform that massively reduced the exposure of black civilians to the NYPD. For the preceding two decades, since William Bratton brought Compstat and intensive policing to New York, the NYPD made tens to hundreds of thousands of stops each year. These stops were legally justified by the Supreme Court's decision in *Terry v. Ohio* (1968), which held that "police can stop a citizen based on founded suspicion that crime may be 'afoot.'" The encounter would proceed to increasing levels of intrusion if suspicion was determined to be credible or reasonable. Reasonable suspicion would permit pointed questioning and frisk or pat down to look for weapons, drugs or other contraband" (Fagan et al., 2009).

Stop, Question, and Frisk (SQF) became a cornerstone of the NYPD's investigative and crime prevention practice, and officers reported organizational pressure to keep their stop rates high (Lerman and Weaver, 2014b; Rayman, 2013). These stops disproportionately targeted black and brown civilians, and especially teenage boys (Gelman, Fagan and Kiss, 2007; Fagan et al., 2009; Goel, Rao and Shroff, 2016). SQF reached a high of nearly 700,000 stops in 2011. That same year, there were 112,115 stops of male black and Latino teenagers between 14 and 18; at the time, about 177,000 New York residents were black and Latino boys between 14 and 18. Even accounting for repeat stops, Fagan et al estimate the probability of being stopped for eighteen- and nineteen-year-old black and Latino males in 2008 at .79 (Fagan et al., 2009). Research has consistently revealed lower success rates (measured by arrest or finding a weapon) for *Terry* stops of black and Latino civilians, and Goel finds that stops of black civilians typically have less *ex ante* justification (Goel, Rao and Shroff, 2016). Taken together, this evidence suggests that the SQF program produced disproportionate exposure to police contact among black and brown civilians in New York.

The SQF program also led to political outcry and legal challenges. In 2008, the Center for Constitutional Rights filed a case challenging Stop, Question, and Frisk as racially biased and unconstitutional. As the case went to trial five years later, in a memo dated March 4, 2013, the plaintiffs requested a procedural change to the "UF-250" forms which officers use to record stops:

“the UF-250 form should be modified to: (i) include a narrative portion for police officers to justify the basis for stops, frisks and searches. . . .” (Mummolo, 2017) The UF-250 form had previously included “the date, time and location of each stop, as well as the reason (suspected crime and other circumstances), suspect attributes and various outcomes such as whether a weapon was found or an arrest was made,” but “critics of SQF had long alleged that this form was insufficient to establish the legality of a stop” (Mummolo, 2017). The NYPD also required officers to “enter details” in their patrol notebooks, but officers were not required to submit these notes and plaintiffs argued that audits of officer notes revealed that they frequently were not recorded.

## **2.1 A Sudden Change to Stop, Question, and Frisk**

In a move that surprised the plaintiffs, the NYPD in fact adopted that reform on the following day. In a memo dated March 5, 2013, Chief of Patrol James P. Hall issued “a new order requiring officers to photocopy and submit these narrative descriptions of the reasons they stopped suspects to supervisors after each shift.” Mummolo finds that this memo led officers to believe that their decisions to implement SQF would be under increased supervision (Mummolo, 2017).

The sudden change in SQF procedure provides a temporal discontinuity in police practice, and an opportunity to observe the consequences of changes in police exposure in the city. After the memo was adopted, police stops fell dramatically and continued to decline over the following year. Mummolo’s paper provides extensive evidence from both plaintiffs and interviews with officers that this intervention surprised officers, and may have been adopted as a strategic move to defang the upcoming court case. This suggests that anticipatory changes by police are not likely to be a confounding factor (Mummolo, 2017).

Mummolo’s interviews with officers confirm that the memo had a significant effect on officers’ perception of how closely their stops would be scrutinized. Despite the high volume of memos in the NYPD, officers told Mummolo that this memo stood out.

“They’re really watching us now,” one officer recalled thinking when the memo was released (Officer 2). Another officer added that before the memo, supervisors, “would

only look at [memo book entries] if someone made an allegation. . . or you had to go to court. . . Now. . . it's basically like they're looking at it. . . without any sort of allegation being made. . . They're trying to find a reason to penalize us," (Officer 1). Supervisors, "obviously look at these things with a fine-tooth comb," said another officer. "We need to protect ourselves, (Officer 3). (Mummolo, 2017)

Before the memo, officers believed themselves to be accountable for delivering a large number of stops (Rayman, 2013). After, they instead believed that stops would need to be carefully justified, and that their supervisors would be proactively monitoring their justifications for stops (Mummolo, 2017).

The memo immediately and substantially reduced in stops; paperwork requirements are a major barrier to police use of particular tactics. Moreover, "the perception of increased risk led some officers to aggressively forego making stops unless they observed something highly incriminating. 'It's forcing people to not get involved in things that otherwise, a few years ago, they would have,'" said one officer (Officer 3)" (Mummolo, 2017). However, it is implausible that the memo changed implicit bias among police officers. The memo thus reduced *exposure* by reducing stops, without changing any confounding features of the NYPD, such as officer's cognitive biases or the criminal and demographic landscape.

How does this reduction in exposure affect racial disparities in police use of force? I find that the probability that force would be used in any specific stop does not change. Using formal tests of equivalence (Hartman and Hidalgo, 2016), I find no significant changes in the rate of use of force at any point within a 30-day bandwidth. However, the number of stops falls substantially at the discontinuity (and continues to decline). Both the daily incidence of use of force on black civilians and the racial disparity in police violence declined significantly.

Could reporting bias explain this effect? Officers reported that they expected their new, more detailed reports to be heavily scrutinized: gone over "with a fine-tooth comb", in one officer's words. Combined with ongoing litigation and the public salience of the issue, this procedural change seems, if anything, more likely to increase reporting of police violence as officers grow more concerned that any omissions might be challenged by supervisors, civilians, or the courts.

## 2.2 Data and Design

This paper examines the consequences of this change in exposure for police violence. In doing so, it overcomes several limitations of existing work on police violence. First, while most of this data draws on inherently sparse records of individual killings, I use NYPD's publicly available Stop, Question, and Frisk data. These records provide continuous, highly granular data on police-civilian interactions. The NYPD SQF data from 2008-2015 describes over 3.2 million stops, including 701,989 instances of use of force by police. Second, this data set provides comprehensive measures of a set of police interactions with their outcomes, avoiding the problem of constructing a reference population. The data cover *all* stops, not just stops which turn out to have force, so they are not skewed by the reverse causality described above.

Prior studies of police violence focus largely on fatal shootings (Fryer Jr, 2016; Ross, 2015). These are inherently rare events, leading to sparseness in the data and difficulty assessing the existence and drivers of variation. In addition, the difference between fatal and non-fatal shootings can be small, perhaps three minutes in the ambulance on the way to the emergency room. This random error in measuring *potentially* lethal events exacerbates the problem of sparseness.

Finally, studies of police violence have rarely used experimental or quasi-experimental techniques, leaving substantial problems with omitted variable bias in the literature. By leveraging a temporal discontinuity, I can compare the same officers in the same city policing the same population, with different levels of exposure. This study is the first to my knowledge to examine the effect of a procedural change that reduces exposure – and disparate exposure – on racial disparities in police violence. Because the data are coded with precise dates, rather than by month or year, the change induced by the stop can be identified more accurately.

## 2.3 Analysis and Results

The outcome of interest is the number of stops in which force is used on a civilian. The incidence of use of force is particularly important because these occasions – not merely the rate – produce negative consequences for individuals, communities, and state legitimacy. Figure 2 shows the

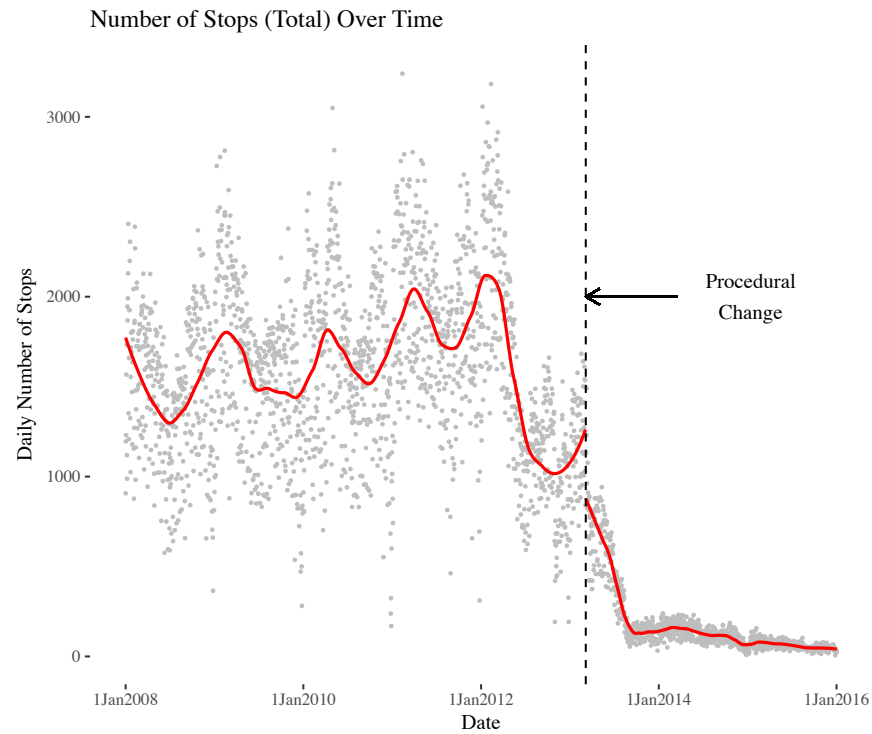
memo's effect on the total number of stops, as well as stops of white and black civilians broken down by race. There is a clear discontinuity at the memo in the number of stops. Dividing the stops by race, though, reveals that this discontinuity results from a decrease in the number of stops of African Americans, not of whites.

The research on the cognitive mechanisms behind implicit bias suggests that officers have a lower threshold for stopping black civilians; analysis of stop, question, and frisk data in New York confirms that officers typically have less evidence *ex ante* when they stop black civilians than when they stop whites (Goel, Rao and Shroff, 2016). As the officer quoted above said, "It's forcing people to not get involved in things that otherwise, a few years ago, they would have" (Mummolo, 2017). When police believe they need more concrete justification for stops, they may forego the stops of precisely the groups of people for whom they previously needed less justification. Regardless of the precise mechanism, this shows a substantial decrease in exposure. Disparate exposure also declined over this period: in the 30 days before the memo, the risk of being stopped for black civilians was 8.14 times that for whites, while in the 30 days post-memo the risk disparity fell to 6.11 – still high, but substantially lower.

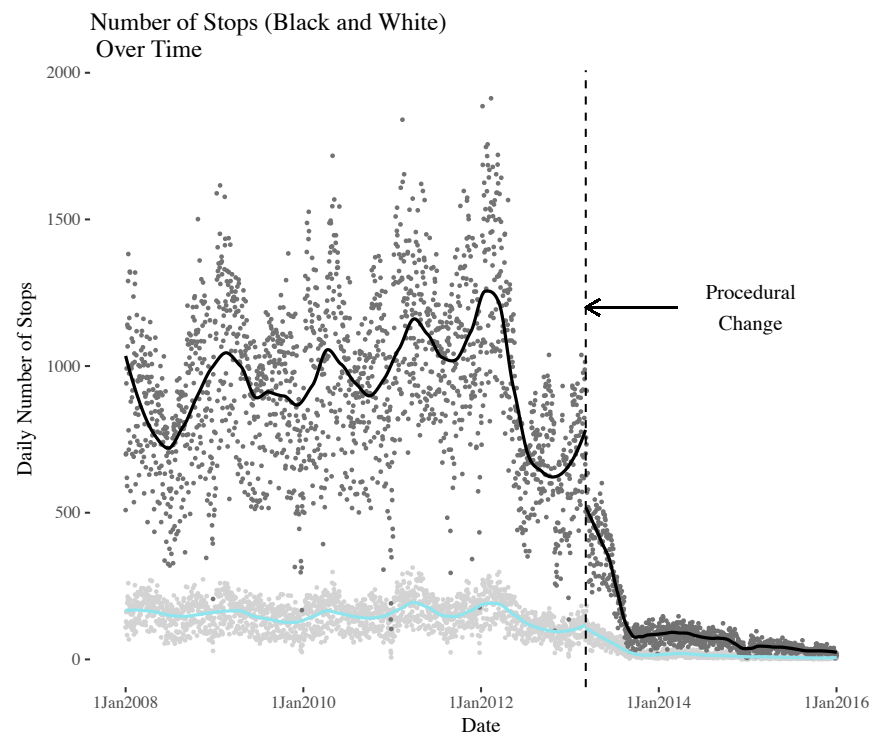
Figure 3 shows the change in the rate of use of force by race. Across the period of the memo, there is no discontinuity in the share of stops which include force for white or black civilians. Over the entire time period, stops of black civilians are between 4 and 9 percentage points more likely to involve force than stops of whites. In Appendix A, I use formal equivalency tests to demonstrate that the rate of use of force did not change when the memo was issued.

Figure 4 shows the change in the total number of stops which included use of force. There is no discontinuous change for white civilians; in contrast, for black civilians, the daily number of incidents with use of force drops precipitously, mirroring the decline in stops almost exactly. The magnitude of this change is substantial.

These visualizations convey the scale of the change in exposure and the accompanying change in use of force on civilians by NYPD, but I also use formal discontinuity tests to estimate the effect of the procedural change on exposure to both stops and use of force for black civilians. I use an



(a)



(b)

Figure 2: This figure shows the daily number of stops from 2008 to 2015. Panel (a) shows all stops (pooled). Panel (b) shows stops for black and white civilians separately. The solid line shows a locally weighted LOESS regression predicting the number of stops, without adjustment for covariates.



Figure 3: This figure shows the percent of stops with use of force for white and black civilians from 2008 to 2015. Here the stop is the unit of analysis/prediction for the LOESS regression.

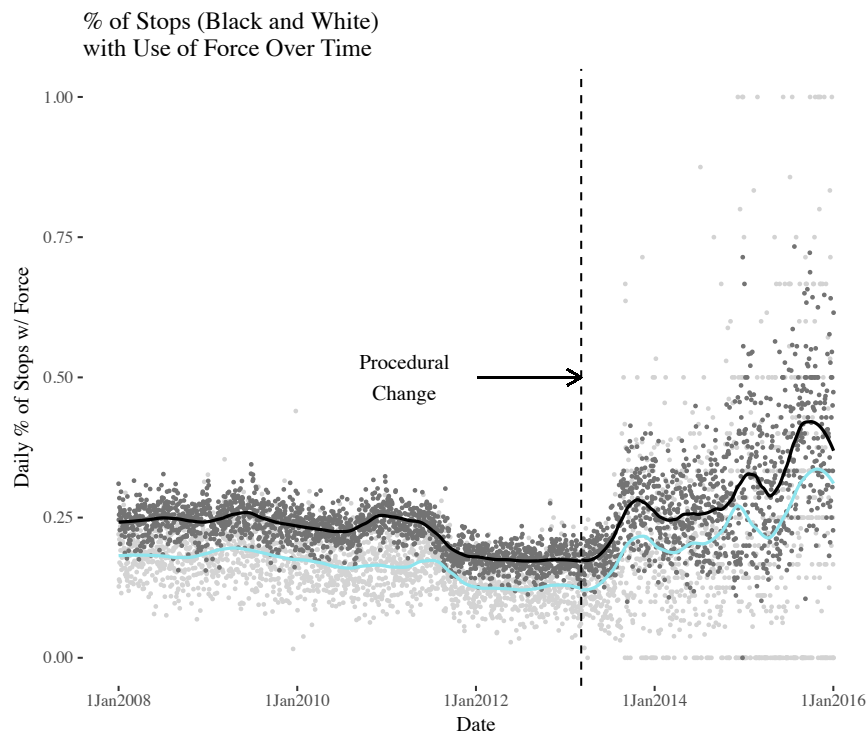
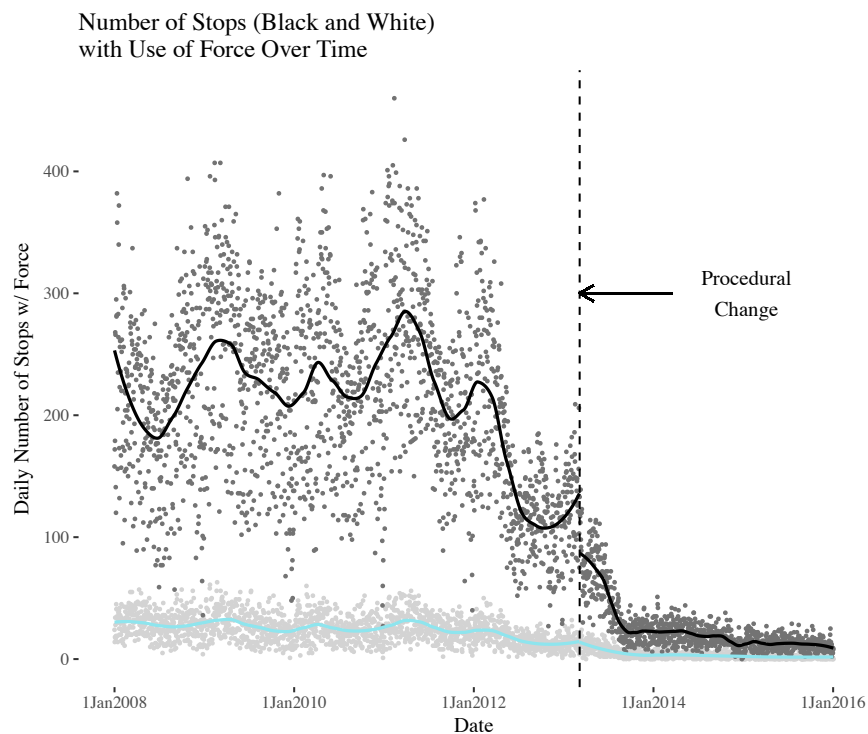


Figure 4: This figure shows the number of stops with use of force for white and black civilians from 2008 to 2015, aggregated by day.



interrupted time series, a form of regression discontinuity where the discontinuity is temporal. The NYPD's SQF dataset provides an ideal environment for this type of analysis: stops are measured continuously and dated to the day, reducing concern about confounding variables – other policies, demographic shifts, or crime-related variables – that might change given a longer period between observations. The well-defined moment of intervention also protects against researcher discretion.

In Appendix A, I assess the rate of use of force before and after the memo. Using formal tests of equivalence, I find no difference in  $p(\text{force}|\text{stop})$  in the pre vs post memo period. Any change in use of force is therefore not the result of changes in the within-stop probability of use of force.

## 2.4 Sharp Regression Discontinuity in Number of Stops

To assess the impact of the memo on disparate exposure and the resulting experience of force, I estimate changes in the number of stops and incidents with use of force by race. I sum data by day and estimate the effect on the number of stops (or incidents with use of force) each day. Because summing by day produces a much smaller number of units, I estimate these quantities using a 100-day bandwidth. Aggregating by day also allows me to include optional controls for year, month, and day of the week. The identifying assumption required to assign the change in outcomes to the memo itself is continuity in potential outcomes: that no other factor which affects the outcomes changed at that precise moment (de la Cuesta and Imai, 2016). Given the granularity of the data and the precise, clearly identified moment of the intervention, these assumptions seem plausible.

$$\text{stops}_d = \alpha + \gamma \text{memo} + \beta_1 \text{day} + \beta_2 \text{day} * \text{memo} + \epsilon_d \quad (1)$$

$$\text{force}_d = \alpha + \gamma \text{memo} + \beta_1 \text{day} + \beta_2 \text{day} * \text{memo} + \epsilon_d \quad (2)$$

Here,  $\text{stops}_d$  is the number of stops in a particular day, while  $\text{force}_d$  is the number of stops in which any type of force was used in the encounter.  $\alpha$  is the intercept;  $\text{memo}$  indicates whether the incident fell before or after the memorandum; and  $\epsilon_d$  is the error term. I estimate both conventional

errors and heteroskedasticity and autocorrelation consistent standard errors.

There are substantive debates about the best strategies to estimate regression discontinuity models (Imbens and Lemieux, 2008; Eggers et al., 2015; Gelman and Imbens, 2014; de la Cuesta and Imai, 2016), and I therefore estimate multiple models. In specifying the function of the running variable, I test local polynomials of orders 0, 1, and 2 for the running variable (i.e. difference in means, linear, and quadratic). Cubic and higher-order estimators have poor properties for estimating regression discontinuity effects (Gelman and Imbens, 2014). All models also include interaction terms between the running variable (day) and the assignment variable (a binary variable indicating whether the day is before or after the memo). In the sharp RDD analyzed in this section,  $\gamma$  captures the effect of being in the post-memo period.

Table 1: OLS Estimates of Change in Daily Stops With Force (“force”) and Daily Stops (“stops”) at Treatment (Memo) Using 100-Day Bandwidth

	Difference in Means	Difference in Means <sup>†</sup>	Linear	Linear <sup>†</sup>	Quadratic	Quadratic <sup>†</sup>
$\Delta$ force (black)	-43.63 (5.835)*	-37.021 (17.775)*	-49.938 (11.917)*	-34.179 (18.252)	-43.781 (19.848)*	-25.525 (15.421)
$\Delta$ stops (black)	-261.61 (39.216)*	-123.508 (58.486)*	-303.011 (66.789)*	-111.593 (61.149)	-265.392 (104.354)*	-94.763 (71.17)
$\Delta$ force (all)	-65.89 (9.299)*	-44.467 (21.498)*	-71.649 (17.633)*	-38.666 (22.189)	-63.684 (29.078)*	-28.037 (22.832)
$\Delta$ stops (all)	-399.37 (67.403)*	-143.649 (73.226)*	-450.857 (109.864)*	-126.335 (74.527)	-413.552 (175.509)*	-107.474 (83.413)
$N$	200	200	200	200	200	200

<sup>†</sup> Includes controls for year, month, day of week, and prior days hit rate.

Maximum of homoscedastic and HAC standard errors in parentheses. \*  $p < 0.05$

In Table 1, all point estimates of the effect on the number of stops and amount of force used – for the total population and for African Americans as a subgroup – are negative, and most are significant. The lowest estimate is 25 avoided uses of force on black civilians per day: an enormous substantive effect considering that there were an average of 98 uses of force per day in the 100-day bandwidth. To assess significance, I calculate both conventional and heteroskedasticity and autocorrelation consistent (HAC) standard errors, then report the larger of the two. In Appendix C, I also calculate these results using the robustness techniques suggested by Calonico et al, including data-driven bandwidth selection, robust, bias-corrected confidence intervals, and a kernel estimator

for improved precision (Calonico, Cattaneo and Titiunik, 2014). I find broadly consistent results.<sup>2</sup>

While the overall differences cannot be attributed to the memo, the longterm decline in Stop, Question, and Frisk had even more massive substantive consequences. In the period before the memo, from January 1, 2008 to March 4, 2013, the NYPD used force on an average of 350 civilians per day, 210 of them black. After the memo, and the concomitant procedural reduction in the use of SQF, the NYPD used force on 39 civilians daily, 24 of them black – an order of magnitude reduction in the number of black civilians who experience use of force from the NYPD. Overall, in the post-memo period, 185 fewer black civilians experienced the use of force by the NYPD *each day*.<sup>3</sup>

## 2.5 Risk Disparities

The result of this procedural change was to reduce civilian exposure to police, and hence to reduce civilian exposure to police violence. The racial disparity in risk, measured by the risk ratio for use of force, also declined: from 11.32 to 8.85 (using a 100-day bandwidth), a 22% reduction.<sup>4</sup> This decline almost exactly mirrored the reduction in racial disparities in exposure, which fell 24% from 8.38 to 6.35.

African Americans' exposure risk remained much higher than that of whites, and the probability of force in any given stop was higher for black pedestrians than for whites both before and after the memo. Still, there is robust evidence that the reduction in disparate exposure was sufficient to massively reduce the exposure of African Americans in New York City to police violence. In the 100 days before the memo, there were 12,061 uses of force on black civilians; in the 100 days after, there were 7,616, nearly 5,000 fewer.

To understand the relative importance of disparate exposure and encounter-level racial inequal-

---

<sup>2</sup>Although the minimum estimate is somewhat lower, the Calonico et al. algorithm uses very large bandwidths (over one year total) to improve precision because the data are aggregated to the day level. The clear effects demonstrated within a 100-day bandwidth are more likely to be the direct effects of the policy change.

<sup>3</sup>In Appendix B, I also use fuzzy regression discontinuity as an alternative strategy to estimate the elasticity of this effect. The estimates are consistently in the range of .16 uses of force per stop.

<sup>4</sup>Results are substantively identical using a 30-day bandwidth. Population data comes from the ACS estimates for New York City in 2013.

ities, I simulate two outcomes in New York. If we kept disparities in exposure at their pre-memo level in SQF but equalized the probability of use of force when police stop black and white civilians, the number of occasions of use of force in the 100 days prior to the memo would have been 8924: a meaningful reduction from the actual number of 12061. If, instead, we equalized exposure rates, but left racial disparities within stops, the number of uses of force on black civilians would fall by an order of magnitude, to a mere 1440 uses of force.

These findings imply that disparate exposure is a major driver of the racial disparities in the experience of police violence.

### **3 Tradeoffs? Broken Windows, Gun Collars, and More**

This paper provides evidence that significantly reducing exposure to policing reduced the exposure of black New Yorkers to police use of force. Frequent police use of force is costly to both individuals and society. Those stopped by police experience substantial costs to their time, dignity, and sense that they can rely on police; when force is used, they suffer both the physical costs of the pain and injuries inflicted by police force, and the costs of alienation from the state. Police use of force also threatens the liberty and safety of individuals who know they risk being stopped by police, and fear the possibility of violence. Moreover, police use of force is costly to society, imposing costs from litigation to community alienation to, potentially, additional violence.

#### **3.1 Gun Collars**

While intensive policing had several goals, one central justification for Stop, Question, and Frisk as a policy was the goal of removing guns from circulation. In 1994, James Q. Wilson, one of the primary architects of broken windows policing, published an article in the *New York Times Magazine* arguing that "the most effective way to reduce illegal gun-carrying is to encourage the police to take guns away from people who carry them without a permit. This means encouraging the police to make street frisks" (Wilson, 1994). Fagan and Davies, drawing on numerous sources,

also identify weapons enforcement as a primary goal of expanded street stops (Fagan and Davies, 2000).

The results, though, suggest that *Terry* stops more often resulted in using force on a civilian than confiscating a weapon. In 2011, at the height of Stop, Question, and Frisk, the NYPD stopped 1087 black civilians daily. Officers used force on an average of 241 black civilians each day, while recovering only 8.5 weapons each day from the black civilians they stopped. Even worse, many ordinary folding knives of the type frequently carried for work in the building trades are misclassified as “gravity knives” under New York law, and treated as weapons (Campbell, 2014). Since guns were the central justification, the most dangerous type of weapon considered, and the most valid measure of true weapon recovered, the ratio of *uses of force* to *guns recovered* is the best measure of the trade-off between weapons enforcement and the use of force on civilians.

In 2011, NYPD recovered 591 guns from black civilians: 1.62 guns per day. A stop of a black civilian was 149 times more likely to end in the use of force than in the recovery of a gun.<sup>5</sup> Table 3 shows the stop outcomes for black civilians by year.

Table 2: Outcomes of Stops of Black Civilians by NYPD, 2008-2015

	Stops	Weapon	Gun	Force	<b>Ratio</b>
2008	305878	2682	616	75643	<b>123</b>
2009	343887	2998	575	86076	<b>150</b>
2010	352029	2970	601	82778	<b>138</b>
2011	396859	3118	591	88077	<b>149</b>
2012	318420	2476	540	56075	<b>104</b>
2013	116108	1527	294	21978	<b>75</b>
2014	26973	588	146	6979	<b>48</b>
2015	13267	481	130	4594	<b>35</b>

It is notable that force has continued to decline with stops. In 2011, 88,077 black civilians experienced the use of force after being stopped under SQF. By 2015, that number had fallen by an order of magnitude. This is a truly major change in civilian risk. Over the same period, the ratio of use of force to guns recovered also fell – consistent with the change in exposure risk, and

<sup>5</sup>This imbalance is even more severe for stops of white civilians, but their vastly lower exposure risk means the consequences for police legitimacy and public well-being are less severe.

Mummolo's finding that weapons recovery remained a high priority (Mummolo, 2017).

Moreover, this finding casts doubt on the substantive importance of the improvement in weapons recovery rates identified by Mummolo. Certainly, Mummolo is right that "in the short term ... this intervention appears to have spared many individuals from being needlessly investigated by police while doing little to impede the recovery of weapons" (Mummolo, 2017). This framing, though, obscures the reality that the NYPD recovered fewer than two guns per day from black civilians in its most active year. (Given the concerns raised above about "gravity knives" and the centrality of gun violence to the discussion of weapons recovery, gun recovery is a much better measure than the more generic weapons recovery.) Despite its importance in justifying SQF, "getting guns off the street" is a vanishingly rare outcome for SQF, while the use of force is a common one.

### 3.2 Intensive Policing and Crime Reduction

Policy debates must also engage the argument for intensive policing. Broken windows advocates argue that SQF and other forms of intensive policing reduce crime by producing an orderly environment that deters more serious crime (Wilson and Kelling, 1982), and that the racial inequalities in disparate exposure are offset by particular benefits in homicide reduction for race-class subjugated communities (Smith and Purtell, 2008). In their expert brief supporting the NYPD in the *Floyd* lawsuit, Smith and Purtell argue that SQF reduced crime – and did so disproportionately in race-class subjugated neighborhoods (*Floyd, et al., v. City of New York, et al., 2017*).

Two things are worth noting in this context. First, there is reason to doubt that SQF in fact reduces crime. Smith and Purtell's findings have faced methodological critiques by researchers who have found the effects of SQF on crime to be minimal (Rosenfeld and Fornango, 2014; Schneiderman, 2013) or modest (Weisburd et al., 2016). Wilson and Kelling's initial argument for Broken Windows policing, which became the central justification for SQF, relied on a particular theory of how crime developed: that public disorder leads to more serious violence (Wilson and Kelling, 1982). More recent studies have found that interpersonal conflict, not public disorder, precedes violence in city neighborhoods (O'Brien and Sampson, 2015). Analyses of exogenous variation in

the number of stops have found little effect on crime (Mummolo, 2017; Chandrasekher, 2016). The decline in stops analyzed in this paper was not accompanied by an increase in homicides, the best measured violent crime (Mummolo, 2017). An analysis of a 1997 NYPD slowdowns motivated by a contract dispute finds no evidence of increases in most serious crimes when police reduce enforcement (Chandrasekher, 2016).

Policing in New York, and to some extent in other cities, has moved away from mass stops and towards hot spot policing, where officers focus on creating a presence – with or without arrests – in areas most affected by crime. However, there are serious concerns with hot spot policing: in particular, Lum and Isaac show that data-driven police deployment generates the appearance of higher crime rates in heavily-policed neighborhoods: because they are present, officers observe crimes that would otherwise go unreported (Lum and Isaac, 2016). Hot spot identification may therefore be unreliable; indeed, both the process of identifying hot spots and the process of expanded deployment may actually produce disparate exposure.

Second, if it *is* the case that SQF reduces crime, or reduces crime disproportionately in race-class subjugated communities, the relationship between *exposure to policing* and *exposure to force* remains relevant. How much less intra-communal violence is needed to justify additional use of force? In New York City in 2013, each stop added .15 uses of force – an important consideration to weigh against benefits in reduced non-police violence. Meanwhile, the probability of recovering a gun during a particular stop has never exceeded .01.

This paper draws data largely from New York City, so these elasticities may be quite different in other cities. Researchers could benefit from substantially more data on police stops and use of force. Nevertheless, this paper provides important context for evaluating the consequences of police intervention.



## 4 Significance

Police use of force matters most centrally because the approximately 1500 people killed by police each year matter, as do the thousands of New Yorkers who experience force at the hands of the NYPD. Use of force matters because of the costs it imposes on the individuals struck, tased, or beaten by police: the physical pain and discomfort, but also the violation of dignity, intrusiveness, and even the time and labor required to file a complaint.

The consequences of police violence, lethal and not, go far beyond these individuals, though. Policing is the main way by which core state services – justice and safety – are delivered. Policing is inevitably an exercise in coercion, but formal justice is a valuable state service. Police resolve disputes about stolen property and public nuisances. They provide a backstop to the civil courts’ enforcement of private legal arrangements: individuals who do not comply with court orders may find police enforcing them. When individuals encounter violence, police intervene to provide safety: enforcing restraining orders, identifying and apprehending those responsible for violence.

Police violence undermines access to these services. Even when committed by a small number of officers, it poisons the well, damaging trust in government and access to state services. People who are afraid of involving police become targets for robbery, because they cannot draw on police protection (Goffman, 2014). This in turn contributes to retaliatory violence, as people defend their physical security or seek resolution to violent actions without access to legal institutions – a dynamic Miller describes as “racialized state failure” (Leovy, 2015; Miller, 2016). The demographic group most likely to be killed by police is the very demographic group most likely to be murdered by civilians: young black men, who make up about 6000 of the 33000 gun deaths in the United States. Mistrust of police, low solve rates for black homicide victims, and other aspects of the carceral state exacerbated by police violence likely contribute to the use of homicide as a tool for dispute resolution and vengeance (when state intervention is lacking) (Leovy, 2015).

The uneven distribution of both police contact and police violence – black Americans are both far more likely to be shot by police and far more likely to be arrested by police than white Americans who engage in the same behaviors – leads to spillover consequences that affect communities

broadly (Ross, 2015; Brame et al., 2012; on the States”, 2009; Rodriguez and Emsellem, 2011; Burch, 2013). The fact of inequality in how different groups are treated shapes individuals’ beliefs about the justice, efficacy, and trustworthiness of the system: even individuals who do not personally observe or experience police violence observe racial differences in how people are treated and conclude that police lack legitimacy.

Finally, police use of force is expensive. Between 2009 and 2014, New York City spent nearly half a billion dollars on police-related settlements, many of them the result of police use of force. Resources spent compensating victims of police use of force could be directed to other public purposes.

How can policymakers reduce the costs of police violence to individuals and to society as a whole, and the racial inequalities in police violence? The standard answer has been to focus on the cognitive biases of police officers, encouraging departments to adopt training protocols to reduce biased behavior *within* interactions. From the perspective of procedural justice, it is critical to ensure that police are treating white and black Americans equally during all police encounters. However, focusing on the cognitive and psychological processes of individual police officers leaves reformers trying to change behavior in high-intensity situations, with little certainty of success.

More importantly, focusing on within-stop behavior and cognitive biases ignores the critical role of disparate exposure in producing racial inequality in police violence. The findings in this paper suggest that inequalities in the steps preceding the decision to use force play a major role in creating inequalities in the experience of police violence. Redressing racial inequality requires a broader perspective.

For the well-being of the families and communities at risk for violence from police, *reduced violence* is as critical as within-stop fairness. Each person who experiences a police stop is at risk for police violence; each person who experiences police violence adds a story to the bone-weary cynicism Shaw describes, and the damage of over-policing and police violence (Shaw et al., 2015). Improving disparities in exposure may not fully address the moral responsibilities of police departments, but this evidence suggests that it will substantively change the daily experience of

black Americans, and lower their risk of violence at the hands of state representatives.

Moreover, this paper offers evidence that bureaucratic changes that reduce disparate exposure can *work*, effectively reducing inequality by changing deployment and engagement decisions; we lack such evidence for interventions targeting cognitive bias in high-intensity contexts. To address inequalities in police use of force, reformers and police departments must come to see stops not as benign, but as carrying the constant risk of escalation against civilians. Policymakers will need to consider the consequences of stops more broadly, but the risk of escalation should play an important role in assessing the costs and benefits of intensive policing. Reducing the number of occasions on which black civilians involuntarily interact with police is likely to substantively reduce black Americans' exposure to police violence.

## References

- Ball, Patrick. 2016. "Violence in Blue: Police Homicides in the United States." *Granta* (134).
- Brame, Robert, Michael G. Turner, Raymond Paternoster and Shawn D. Bushway. 2012. "Cumulative prevalence of arrest from ages 8 to 23 in a national sample." *Pediatrics* 129(1):21–27.
- Brayne, Sarah. 2014. "Surveillance and System Avoidance Criminal Justice Contact and Institutional Attachment." *American Sociological Review* p. 0003122414530398.
- Burch, Traci. 2013. *Trading democracy for justice: Criminal convictions and the decline of neighborhood political participation*. University of Chicago Press.
- Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6):2295–2326.
- Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2015. "rdrrobust: An r package for robust nonparametric inference in regression-discontinuity designs." *R Journal* 7(1):38–51.

- Campbell, Jon. 2014. "How a '50s-Era New York Knife Law Has Landed Thousands in Jail." *Village Voice* .
- Chandrasekher, Andrea Cann. 2016. "The Effect of Police Slowdowns on Crime." *American Law and Economics Review* 18(2):385–437.
- Christie, Gayre, Simon Petrie and Perri Timmins. 1996. "The effect of police education, training and socialisation on conservative attitudes." *Australian & New Zealand Journal of Criminology* 29(3):299–314.
- Correll, Joshua, Bernadette Park, Charles M Judd, Bernd Wittenbrink, Melody S Sadler and Tracie Keese. 2007. "Across the thin blue line: police officers and racial bias in the decision to shoot." *Journal of personality and social psychology* 92(6):1006.
- de la Cuesta, Brandon and Kosuke Imai. 2016. "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19:375–396.
- Eberhardt, Jennifer L, Phillip Atiba Goff, Valerie J Purdie and Paul G Davies. 2004. "Seeing black: race, crime, and visual processing." *Journal of personality and social psychology* 87(6):876.
- Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall and James M. Snyder. 2015. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races." *American Journal of Political Science* 59(1):259–274.
- Epp, C. R., S. Maynard-Moody and D. P. Haider-Markel. 2014. *Pulled Over: How Police Stops Define Race and Citizenship*. Chicago Series in Law and Society University of Chicago Press.
- Fagan, J., A. Geller, G. Davies and V. West. 2009. "Street stops and broken windows revisited: The demography and logic of proactive policing in a safe and changing city." *RACE, ETHNICITY, AND POLICING: NEW AND ESSENTIAL READINGS*, Stephen K. Rice and Michael D. White, eds., New York University Press pp. 09–203.

Fagan, Jeffrey and Garth Davies. 2000. "Street stops and broken windows: Terry, race, and disorder in New York City." *Fordham Urb. LJ* 28:457.

*Floyd, et al., v. City of New York, et al.,*. 2017. Vol. 959.

Fryer Jr, Roland G. 2016. *An Empirical Analysis of Racial Differences in Police Use of Force* .

Gelman, A., J. Fagan and A. Kiss. 2007. "An analysis of the New York City Police Department's stop-and-frisk policy in the context of claims of racial bias." *Journal of the American Statistical Association* 102(479):813–823.

Gelman, Andrew and Guido Imbens. 2014. "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs." *National Bureau of Economic Research Working Paper Series* .

Goel, Sharad, Justin M Rao and Ravi Shroff. 2016. "Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy." *Annals of Applied Statistics* 10(1):365–394.

Goffman, Alice. 2014. *On the run : fugitive life in an American city*. University of Chicago Press.

Greenwald, Anthony G, Mark A Oakes and Hunter G Hoffman. 2003. "Targets of discrimination: Effects of race on responses to weapons holders." *Journal of Experimental Social Psychology* 39(4):399–405.

Harris, Alexes. 2016. *A pound of flesh: Monetary sanctions as punishment for the poor*. Russell Sage Foundation.

Hartman, Erin and F. Daniel Hidalgo. 2016. "What's the alternative?: An equivalence approach to balance and placebo tests." *Working Paper* .

Imbens, Guido W and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of econometrics* 142(2):615–635.

- James, Lois, Stephen M James and Bryan J Vila. 2016. "The reverse racism effect." *Criminology & Public Policy* 15(2):457–479.
- LaFraniere, Sharon and Mitch Smith. 2016. "Philando Castile Was Pulled Over 49 Times in 13 Years, Often for Minor Infractions." *The New York Times* .
- Leovy, Jill. 2015. *Ghettoside*. Spiegel & Grau.
- Lerman, Amy E. and Vesla M. Weaver. 2014a. *Arresting Citizenship: The Democratic Consequences of American Crime Control*. University of Chicago Press.
- Lerman, Amy E. and Vesla M. Weaver. 2014b. "Staying out of sight? Concentrated policing and local political action." *The ANNALS of the American Academy of Political and Social Science* 651(1):202–219.
- Lum, Kristian and William Isaac. 2016. "To predict and serve?" *Significance* 13(5):14–19.
- Miller, Lisa Lynn. 2016. *The Myth of Mob Rule: Violent Crime and Democratic Politics*. Oxford University Press.
- Moskos, Peter. 2009. *Cop in the hood: my year policing Baltimore's eastern district*. Princeton University Press.
- Mummolo, Jonathan. 2017. "Modern Police Tactics, Police-Citizen Interactions and the Prospects for Reform." *Journal of Politics* .
- O'Brien, Daniel Tumminelli and Robert J. Sampson. 2015. "Public and Private Spheres of Neighborhood Disorder." *Journal of Research in Crime and Delinquency* 52(4):486–510.
- on the States", "Pew Center. 2009. Washington, DC: The Pew Charitable Trusts.
- Paluck, Elizabeth Levy and Donald P Green. 2009. "Prejudice reduction: What works? A review and assessment of research and practice." *Annual review of psychology* 60:339–367.

- Peralta, Eyder and Cheryl Corley. 2016. "The Driving Life And Death Of Philando Castile." *National Public Radio* .
- Plant, E Ashby and B Michelle Peruche. 2005. "The consequences of race for police officers' responses to criminal suspects." *Psychological Science* 16(3):180–183.
- Rayman, Graham A. 2013. *The NYPD Tapes: A Shocking Story of Cops, Cover-ups, and Courage*. Macmillan.
- Rodriguez, Michelle Natividad and Maurice Emsellem. 2011. "65 Million 'Need Not Apply': The Case for Reforming Criminal Background Checks for Employment." *New York: National Employment Law Foundation* .
- Rosenfeld, Richard and Robert Fornango. 2014. "The impact of police stops on precinct robbery and burglary rates in New York City, 2003-2010." *Justice Quarterly* 31(1):96–122.
- Ross, Cody T. 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):e0141854.
- Roussell, Aaron, Kathryn Henne, Karen S Glover and Dale Willits. 2017. "Impossibility of a Reverse Racism Effect." *Criminology & Public Policy* .
- Saltzstein, Grace Hall. 1989. "Black mayors and police policies." *The Journal of Politics* 51(03):525–544.
- Schneiderman, Eric T. 2013. *A Report on Arrests Arising from the New York City Police Department's Stop-and-frisk Practices*. New York State: Office of the NYS Attorney General, Civil Rights Bureau.
- Shaw, Theodore M et al. 2015. *The Ferguson Report: Department of Justice Investigation of the Ferguson Police Department*. The New Press.

- Sim, Jessica J, Joshua Correll and Melody S Sadler. 2013. "Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot." *Personality and social psychology bulletin* 39(3):291–304.
- Smith, Dennis C and Robert Purtell. 2008. "Does stop and frisk stop crime?" *Annual Research Conference of the Association of Public Policy and Management* .
- Soss, Joe and Vesla Weaver. 2017. "Learning from Ferguson: Policing, Race, and Class in American Politics." *Annual Review of Political Science* .
- Weaver, V. M. and A. E. Lerman. 2010. "Political consequences of the carceral state." *American Political Science Review* 104(4):817–833.
- Weisburd, David, Alese Wooditch, Sarit Weisburd and Sue-Ming Yang. 2016. "Do Stop, Question, and Frisk Practices Deter Crime?" *Criminology & public policy* 15(1):31–56.
- White, Ariel. 2015. "Misdemeanor Disenfranchisement? The demobilizing effects of brief jail spells on potential voters." *Harvard Scholar* .
- Wilson, James Q. 1978. *Varieties of police behavior*. Harvard University Press.
- Wilson, James Q. 1994. "Just Take Away Their Guns." *The New York Times* .
- Wilson, James Q and George L Kelling. 1982. "Broken windows." *Atlantic monthly* 249(3):29–38.
- Zimring, Franklin E. 2017. *When Police Kill*. Harvard University Press.



## A Equivalence in $p(\text{force}|\text{stop})$

To estimate changes in the rate of use of force, I estimate the following equation:

$$\text{force}_i = \alpha + \beta_1 \text{memo}_i + \beta_2 \text{memo}_i d_i + \epsilon_i \quad (3)$$

Here,  $\text{force}_i$  is the  $(0, 1)$  assessment of whether any type of force was used in the encounter.  $\alpha$  is the intercept,  $\text{memo}_i$  indicates whether the incident fell before or after the memorandum,  $s_j(d_i)$  is a function of the days, and  $\epsilon_i$  is the error term. I estimate both conventional errors and heteroskedasticity and autocorrelation consistent standard errors.

There are substantive debates about the best strategies to estimate regression discontinuity models (Imbens and Lemieux, 2008; Eggers et al., 2015; Gelman and Imbens, 2014; de la Cuesta and Imai, 2016), and I therefore estimate multiple models, including difference in means as well as linear, quadratic, and cubic time trends with and without controls. The identifying assumption required to assign the change in outcomes to the memo itself is continuity in potential outcomes: that no other factor which affects the outcomes changed at that precise moment (de la Cuesta and Imai, 2016).

I estimate the change in rate of use of force at a variety of bandwidths up to 100 days, and separately for the data as a whole and for black civilians. This 100-day bandwidth covers a total of 182,279 stops over 200 days (111,523 stops of black civilians), including 31,421 uses of force, 19,759 on black civilians and 2,397 on white civilians. Figure 5 shows the point estimates for these tests for black civilians, while Figure 6 shows the estimates for all civilians.

Absence of statistically significant differences does not, per se, demonstrate equivalence. Following Hartman and Hidalgo's recommendations, I use 0.2 times the pooled standard deviation in rates of use of force as an equivalence threshold (Hartman and Hidalgo, 2016). The vast majority of outcomes fall within the equivalence range; in 600 estimates, only one is (very marginally) below the equivalence range. This suggests that we can establish that rates of use of force in the post-memo period, conditional on stops, were equal to or higher than rates of use of force before

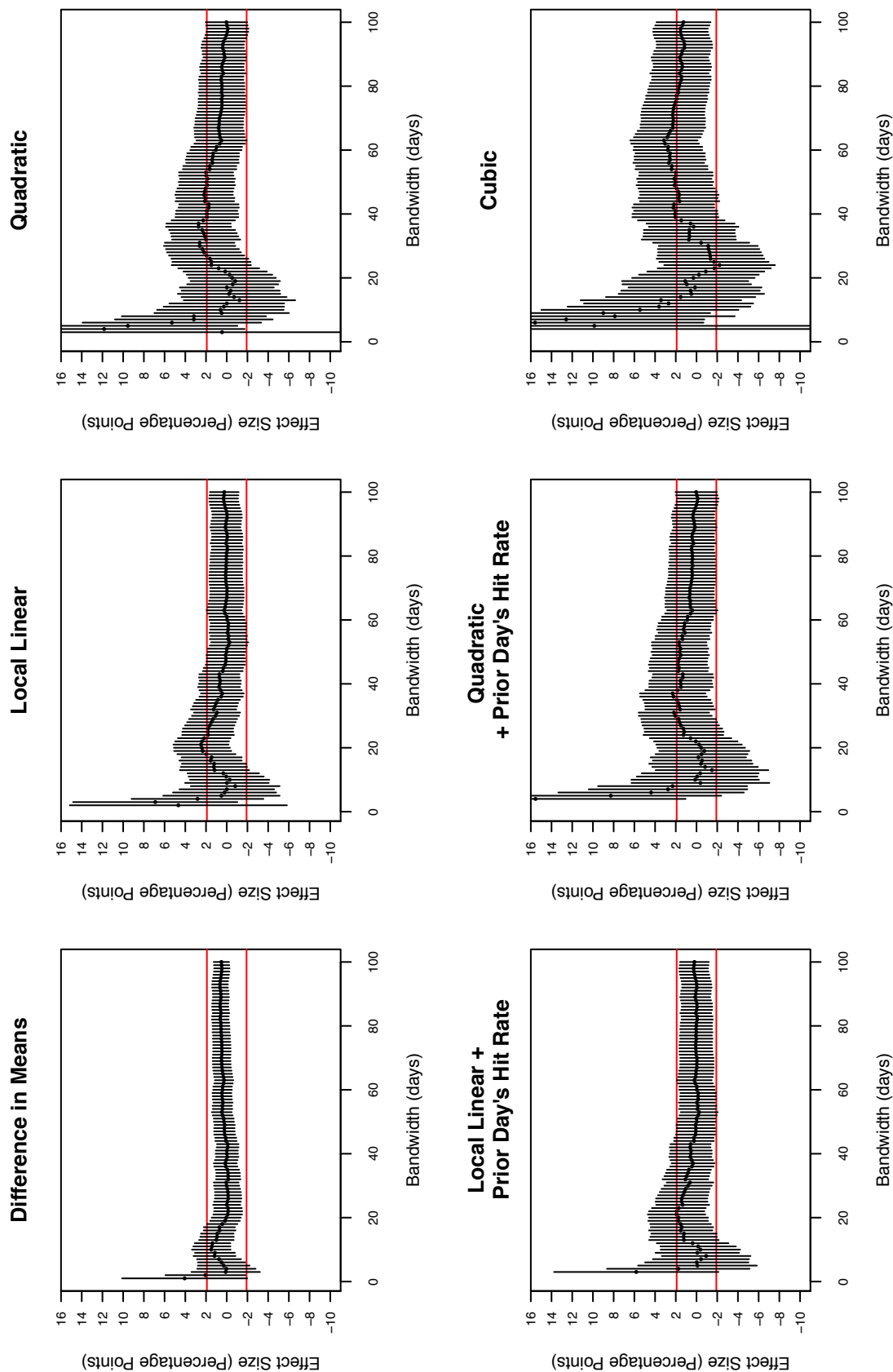


Figure 5: Stops of Black Civilians: Estimates of the change in probability of using force on the day of the intervention using various model specifications and bandwidths. Vertical lines show 95% confidence intervals for each estimate. Horizontal lines show the equivalence range.

the memo was issued.

### A.1 Fuzzy Regression Discontinuity Design: Relationship Between Stops and Uses of Force

To estimate the elasticity of this effect, I use the fuzzy variant of regression discontinuity. In the fuzzy regression discontinuity design, the probability of receiving treatment jumps discontinuously at a defined threshold, but not necessarily from zero to one. This analytical strategy defines the treatment not as the memo, which is the treatment applied to officers in the previous analysis, but as the *stop* itself. Thus, the probability of being treated (stopped) falls discontinuously for black New Yorkers. Using fuzzy RDD, I can estimate the effect of one foregone stop on the number of occasions of use of force. Fuzzy RDD adds the assumptions required for an instrumental variables design to those of a regression discontinuity design: monotonicity, relevance, and exclusion. The graph above supports all three: the number of stops (probability of treatment) strictly declines across the discontinuity, and the change is large, negative, and statistically significant. Because of the granular nature of the data it is unlikely that something other than the memo changed in the NYPD at that time. Finally, the results shown in Figures 6 and 5 suggest that the probability of force in any given stop did not change over time. Within-stop factors leading to use of force did not change with the memo, suggesting that the memo affected use of force only through its effect on exposure, and thus that the exclusion restriction is satisfied.

The first stage equation is

$$stops_d = \alpha + \gamma memo + s_j(d) + \eta_d \quad (4)$$

where  $stops_d$  is the number of stops on that day (i.e. the number of units assigned to treatment),  $\gamma$  is the effect of the memo on the number of stops,  $s_j(d)$  is a function of the forcing variable (date) that can take a variety of forms, and  $\eta_d$  is the error term.

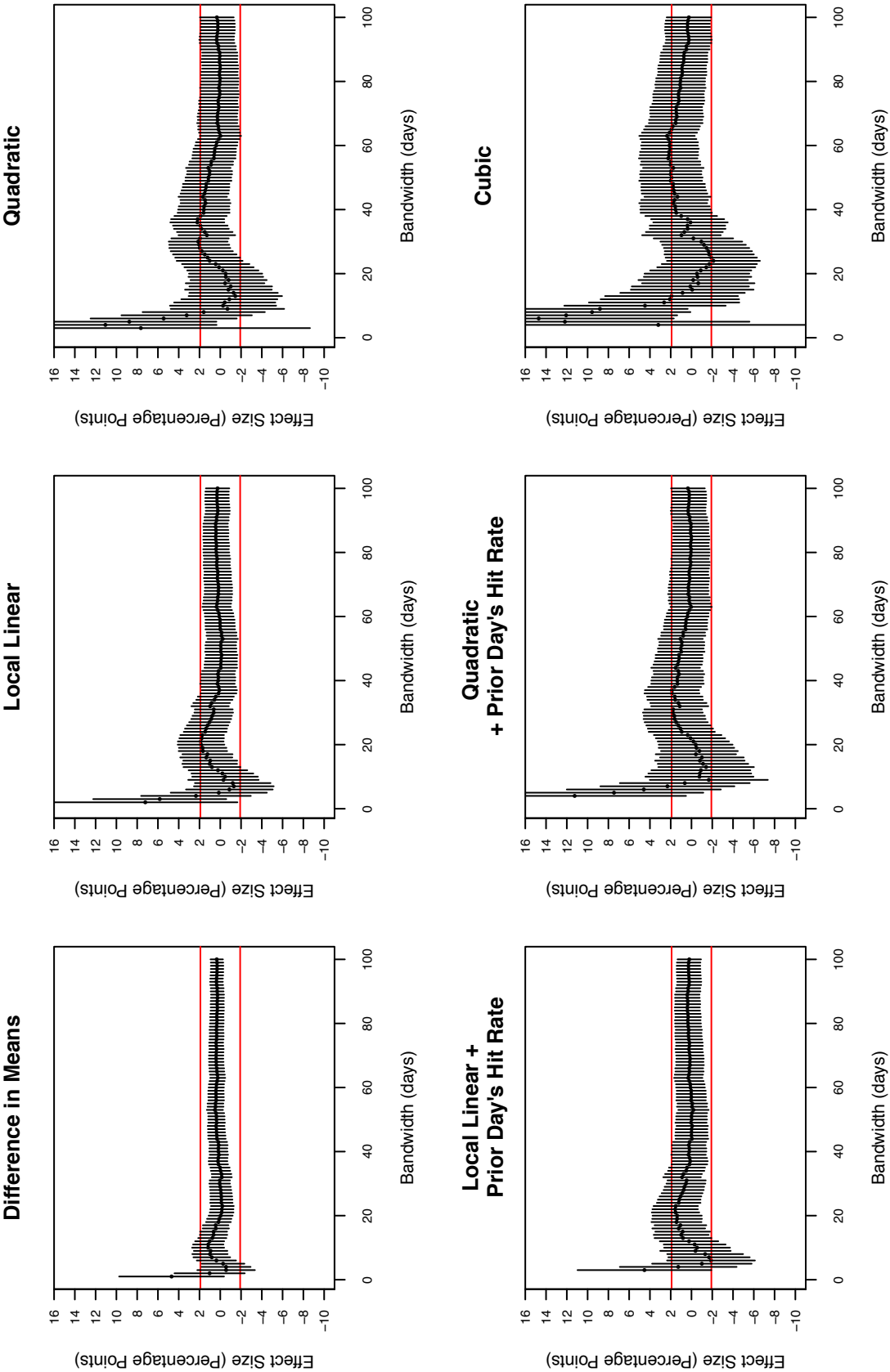


Figure 6: All Stops: Estimates of the change in probability of using force on the day of the intervention using various model specifications and bandwidths. Vertical lines show 95% confidence intervals for each estimate. Horizontal lines show the equivalence range.

The second stage equation is

$$force_d = \alpha + \widehat{\beta stops_d} + s_j(d) + \epsilon_d \quad (5)$$

Table 2 shows the results of this estimate using the 100-day bandwidth, while Table 5 shows the results for the data-driven bandwidths supported by Calonico et al (Calonico, Cattaneo and Titiunik, 2015). The estimates are consistent across various specifications, ranging from .15 to .17. The implication is that eliminating a single stop of a black civilian in this period eliminates between .15 and .17 uses of force, a figure quite consistent with the percent of occasions on which force was used during that time.

Table 3: OLS Estimates of Effect of Changes in Stops on Force Using 100-Day Bandwidth (Fuzzy RDD)

	Difference in Means	Difference in Means <sup>†</sup>	Linear	Linear <sup>†</sup>	Quadratic	Quadratic <sup>†</sup>
$\Delta$ force/stop	0.165	0.168	0.157	0.167	0.155	0.165
(all)	(0.007)*	(0.006)*	(0.008)*	(0.006)*	(0.008)*	(0.006)*
$\Delta$ force/stop	0.167	0.176	0.16	0.176	0.159	0.174
(black)	(0.008)*	(0.007)*	(0.009)*	(0.007)*	(0.009)*	(0.007)*
<i>N</i>	200	200	200	200	200	200

<sup>†</sup> Includes controls for year, month, day of week, and prior days hit rate.

Maximum of homoscedastic and HAC standard errors in parentheses. \*  $p < 0.05$

Again, I re-estimate results using the procedures recommended by Calonico, et al: data-driven bandwidths, robust bias-corrected confidence intervals, and a kernel estimator to improve fit. These results (available in Appendix C) are consistent with those reported here.

## B Robust Estimation

For both regression discontinuity estimates, I use `rdrobust` to follow data-driven procedures to select the ideal bandwidth (Calonico, Cattaneo and Titiunik, 2015). Following Calonico et al, I also use robust, bias-corrected confidence interval techniques in these estimates, which are “not only valid when the usual bandwidth conditions are satisfied (being asymptotically equivalent to

the conventional confidence intervals in this case), but also continue to offer correct coverage rates in large samples even when the conventional confidence intervals do not” (Calonico, Cattaneo and Titiunik, 2014). For most specifications, the data-driven procedure selects a substantially larger bandwidth than 100 days; the 100-day estimates thus, if anything, underestimate the precision with which the effect is identified, and would make the analysis more likely to incorrectly reject the null hypothesis (Calonico, Cattaneo and Titiunik, 2014). Table 4 shows the results using the data-driven bandwidths and robust, bias-corrected confidence intervals. These robust results are also estimated with a triangular kernel, to improve fit.

Table 4: Estimates of Change in Daily Stops With Force (“force”) and Daily Stops (“stops”) at Treatment (Memo) Using Data-Driven Bandwidth

	Difference in Means	Linear	Linear <sup>†</sup>	Quadratic	Quadratic <sup>†</sup>
$\Delta$ force (black)	-42.896	-50.585	-15.725	-52.188	-20.391
	(9.911)*	(9.96)*	(6.974)*	(8.928)*	(7.461)*
N	75	187	435	489	711
$\Delta$ stops (black)	-247.636	-271.919	-63.94	-344.522	-101.717
	(48.431)*	(50.551)*	(33.612)*	(42.773)*	(32.308)*
N	67	155	396	432	806
$\Delta$ force (all)	-62.055	-72.002	-19.897	-75.097	-26.047
	(13.149)*	(13.487)*	(10.335)*	(12.053)*	(10.541)*
N	75	178	373	483	697
$\Delta$ stops (all)	-375.107	-403.345	-83.667	-496.884	-150.423
	(71.271)*	(75.636)*	(52.234)*	(62.427)*	(47.949)*
N	69	151	381	462	846

<sup>†</sup> Includes controls for year, month, day of week, and prior days hit rate.

\*  $p < 0.05$

Table 5 reports `rdrobust` estimates for the fuzzy regression discontinuity. Results are consistent with the estimates using a 100-day bandwidth.

Table 5: OLS Estimates of Effect of Changes in Stops on Force Using Data-Driven Bandwidth (Fuzzy RDD)

	Difference in Means	Linear	Linear <sup>†</sup>	Quadratic	Quadratic <sup>†</sup>
$\Delta$ force/stop	0.161	0.17	0.15	0.152	0.184
(all)	(0.014)*	(0.015)*	(0.029)*	(0.013)*	(0.033)*
N	76	409	314	432	745
$\Delta$ force/stop	0.164	0.176	0.173	0.162	0.177
(black)	(0.017)*	(0.015)*	(0.035)*	(0.015)*	(0.041)*
N	73	393	362	476	663

<sup>†</sup> Includes controls for year, month, day of week, and prior days hit rate.

Maximum of homoscedastic and HAC standard errors in parentheses. \*  $p < 0.05$