

Are Police Racially Biased in the Decision to Shoot?*

Tom S. Clark[†] Elisha Cohen[‡] Adam Glynn[§] Michael Leo Owens[§]
Anna Gunderson[¶] Kaylyn Jackson Schiff[‡]

June 2, 2020

Abstract

Racial disparities in police use of force are common in the United States and may be widest for Black Americans relative to other racial groups. Determining racial bias as the cause of such disparities is difficult due to a lack of systematic use of force data and inferential challenges to discovering and estimating racial bias by police officers. Our theoretical model predicts that in less dangerous situations, police will be more likely to use force against Black civilians than against White civilians. The model implies different fatality rates among White and Black civilians shot by the police, which we empirically evaluate with original data for all officer-involved shootings of civilians from 2005-2017 in nine local police jurisdictions in the U.S. The model and test permit us to begin to credibly assess civilian fatality rates, conditional upon civilians being shot by the police. Our theoretical implication and empirical findings provide novel evidence of racial bias in fatal police shootings of civilians.

*We thank Chuck Cameron, Anna Harvey, Hakeem Jefferson, John Kestellec, Beatriz Magaloni-Kerpel, Pablo Montagnes, John Patty, Zac Peskowitz, Dan Thompson, Georg Vanberg, and Kirsten Widner for helpful comments. Earlier versions of this paper were presented at the 2019 meeting of the American Political Science Association, the Politics of Policing Conference at Princeton University, and seminars at Emory University.

[†]Charles Howard Candler Professor of Political Science, Emory University

[‡]Ph.D. Student, Emory University

[§]Associate Professor of Political Science, Emory University

[¶]Assistant Professor, Louisiana State University

1 Introduction

Political scientists have recently demonstrated considerable interest in effects of contact with the criminal justice system on citizen participation in political life (e.g., White, 2019; Bateson, 2012; Weaver and Lerman, 2010). Perhaps the single most common way in which a citizen comes into contact with the criminal justice system is through the police, who are a near-omnipresent representation of the state and exercise a great degree of discretion when interacting with the public. Moreover, the police are different from most other government representatives in that interactions with them almost always have the potential to turn violent and even lethal. While police use of force is greater in some nations than others, police shootings of civilians are more common in the United States relative to other advanced, liberal democracies (Zimring, 2017). Furthermore, contemporary police shootings of civilians of color raise concerns about racial bias influencing officers' discretion to shoot some civilians but not others. The concern stems partly from the fraught history and contemporary realities of racial disparities in policing, including surveillance, involuntary contact via stop-and-frisk, arrest, and jailing (Brown, 2019; Soss and Weaver, 2017), as well as racial disparities in the use of force by police against civilians of color. However, whether and how often racial bias produces the racial disparities in policing remains an academic and civic puzzle.

It is empirically difficult to discern how much racial bias accounts for racial disparities in police shootings of civilians in the U.S. It is unclear how many police shootings of Black Americans result from their disproportionate contact with police, versus the disproportionate use of force against them by police, versus racial bias by the police (e.g., Knox, Lowe, and Mummolo, forthcoming; Knowles, Persico, and Todd, 2001; Fryer Jr., 2016). Additionally, determining whether racial disparities mainly result from police officers, both White and non-White, demonstrating a differential valuing of the lives of racial minorities, particularly Black lives, in comparison to the lives of Whites is hard without attitudinal data from police officers or complaint data about them. Also, neither police departments nor agencies overseeing them (e.g., U.S. Department of Justice) necessarily track or report all lethal and non-lethal police shootings of civilians, especially by race (Zimring, 2017). Consequently, depending on data, measures, and methods, studies draw contradictory conclusions, ranging from significant differences in the likelihood and speed of shooting Black civilians compared to other civilians (Mekawi and Bresin, 2015) to no racial disparities in fatal officer-involved shootings of civilians (Johnson et al., 2019). In short, even when relatively good data are available, allowing social scientists to observe and describe racial patterns in policing, scholarly consensus on whether police discriminate by race of civilian when using lethal force, let alone nonlethal force, remains elusive.

To better assess whether there is evidence of racial bias in the use of force by police against civilians, measured by shootings, lethal and non-lethal, we develop a model of civilian-police encounters that yields

empirical implications for evaluating racial bias in officer-involved shootings (OIS) data. In our model, informed by studies of the transactional nature and iterative process of police-civilian encounters (Binder and Scharf, 1980; Terrill, 2005; Kahn et al., 2017), civilians and police make choices about their behavior. That behavior may include actions that escalate their encounters towards harm, including police shootings of civilians (and civilian violence against police). Our model predicts that *if* officers are racially biased (or perceived as biased) in favor of shooting Black civilians, Black civilians should systematically behave differently when interacting with the police. Specifically, Black civilians should choose actions intended to reduce police officers perceiving them as threatening to officers relative to how White civilians behave. In other words, in less dangerous confrontations between police and civilians, police officers will be more likely to use force against Black civilians than against White civilians. Consequently, police shootings of Black civilians should result in more non-fatalities than fatalities, as we will empirically show.

Although current methodological debates center on the complex selection processes that affect civilian contact with the police, which complicate the ability of social scientists to draw inferences about racial bias from observed racial patterns even among those individuals who interact with the police (Johnson et al., 2019; Knox and Mummolo, 2019; Knox, Lowe, and Mummolo, forthcoming), our approach seeks to side-step that selection process by examining trace evidence of discrimination that exists after any such selection process plays out. To test our model, we employ a “hit rate” test that permits evaluating evidence for racial discrimination, which is a method other social scientists leverage for discerning racial bias across a range of discretionary decisions, including mortgage lending, bail-bonding, editorial acceptances, hiring, capital sentencing, and police traffic stops (Canner, Gabriel, and Woolley, 1991; Munnell et al., 1992; Ayres, 2002; Knowles, Persico, and Todd, 2001; Persico and Todd, 2006; Alesina and La Ferrara, 2014).

We test that implication with novel data covering all OIS in nine US localities, which we obtained through public records requests. Our data include all instances of civilians shot by local police—both fatally and non-fatally—and the race of civilians, along with other attributes of the police-civilian encounters. Consistent with our theoretical expectation, we find that Black civilians are more likely to survive an OIS, reflecting, we posit, a higher degree of racial bias in the decisions by officers to shoot Black civilians compared to non-Black civilians. Additionally, we estimate a lower bound on the magnitude of racial bias in the decision to shoot a civilian, guided by Knox, Lowe, and Mummolo, forthcoming. Borrowing their technique, we conceptually divide Black civilians that were shot into Black civilians that would have been shot had they been White and Black civilians that would not have been shot had they been White. The *proportional* size of the second group is our parameter of racial bias. To estimate a lower bound for this quantity, we evaluate the difference in fatality rates of White and Black civilians shot by the police in the nine localities relative to the White fatality rate in them, where we posit fatal shootings are more likely to be justified as “reasonable” shootings

from the perspective of police departments, and that non-fatal shootings are more prevalent among Black civilians compared to other groups.¹

Our theory and findings demonstrate that identifying racial bias in police decision-making is possible, buttressing other research (Knowles, Persico, and Todd, 2001; Persico and Todd, 2006; Knox and Mummolo, 2019; Knox, Lowe, and Mummolo, forthcoming). That alone is important in light of the continuing need to understand discretion by the police as “street-level bureaucrats” and how much race affects policing, including use force and its severity. Plus, our theory and findings about the most extreme form of police use of force bear on classic concerns in political science, including but not limited to the exercise of power by the state, democratic accountability, and equality under the law.

2 Police Discretion in Use of Force

Encounters with the police are among the most common encounters civilians have with government agents (Jacob, 1972; Soss and Weaver, 2017), though a key contrast with other civilian encounters with government agents is that contact with police officers has the potential for violence. Much research has focused on how officers exercise their discretion to use force (e.g., Terrill, 2011) and the potential for racial disparities discretion can cause.

2.1 Racial Disparities in Police Use of Force

Generally, the expectation is police are more likely to use force and more of it against Black civilians than against White civilians (James, Vila, and Daratha, 2013; Goff et al., 2016; Jetelina et al., 2017). Whether police use their discretion to employ more force against Black civilians than White civilians is well-studied but experimentally and observationally, often finding that officers are more willing to use force against Black civilians than against White civilians (Correll, Park, Judd, Wittenbrink, et al., 2007; Mekawi and Bresin, 2015; Eberhardt et al., 2004; Buehler, 2017; Sikora and Mulvihill, 2002; Johnson et al., 2019; Worden, 2015; Engel and Calnon, 2004; Schuck, 2004; Terrill, 2005; Baumgartner, Epp, and Shoub, 2018). Furthermore, the recent availability of “big data” on incident-level public data on police-civilian encounters (e.g., New York City’s Stop, Question, and Frisk program) has enabled social science applications of powerful analytic tools to deepen evidence of racial disparities in police use of force, and treatment of civilians more generally (e.g., Fryer Jr., 2016; Voigt et al., 2017; Pierson et al., 2017; Gelman, Fagan, and Kiss, 2007; Goel, Rao, and Shroff, 2016; Mummolo, 2018).

¹Although we lack data on all instances in which police draw their weapons, inclusive of pulling guns but not firing them, the use of such data, if available, would likely produce a larger estimate of the lower bound (Worrall et al., 2018). This will be discussed in detail below.

However, some studies temper or contradict claims and the expectation of racial bias in police use of force, particularly shootings (e.g., Worrall et al., 2018), suggesting racial bias in policing may not necessarily increase the likelihood of use of force against Black civilians. Some evidence, drawn from studies that are typically observational and limited by concerns about unmeasured confounding and/or misapplied methods (J. H. Garner, Schade, et al., 1995; J. Garner and C. Maxwell, 1999; J. H. Garner, C. D. Maxwell, and Heraux, 2002; Alpert and Dunham, 2004; Fryer Jr., 2016; Johnson et al., 2019), suggests we should expect and observe either no or smaller-scale racial disparities in whether or how much the police use force (e.g., shootings) against civilians. Plus, a “counter bias” may exist, inducing officers to be extra sensitive to the potential negative consequences of using force against racialized civilians (James, Vila, and Daratha, 2013). The negative consequences of using force and more of it against Black civilians might be *higher*, not lower, than they are for using force against White civilians. (However, the strength of evidence for such an effect is debatable (Johnson et al., 2019; Knox and Mummolo, 2019).)

2.2 Challenges to Inferring Racial Bias

“In the police shooting context, there is a concern that officers, despite their best intentions and/or conscious beliefs, will subconsciously let preconceived ideas about certain individuals influence their decision processes” (Worrall et al., 2018, p. 1176). This includes their racial beliefs, which may bias their behaviors during police-civilian encounters. However, inferring racial bias is challenging. First, there can be different conceptions of racial bias. The quote above focuses on the bias of the officer that shoots, while one might alternatively focus on the bias of the officer (or algorithm) that makes deployment decisions. In this paper, we focus on the former type of “officer on the street” bias, while acknowledging the importance of the latter type of “command center” bias. Second, the “race” of an individual is not randomly realized during police encounters with civilians.² As a consequence, any inference about the causal effect of the race of a civilian on police use of force, or other behaviors (e.g., driver or pedestrian stops) depends on the comparability of incidents. Unfortunately, confounds in the use of force can be difficult to measure. Even if one can account for the lack of observed outcomes for officer-civilian encounters that never take place, empirical tests for racial bias still must account for confounding factors that affect contact and the use of force (Knox, Lowe, and Mummolo, forthcoming). Race, for example, may be correlated with other characteristics (e.g., income, education, geography, employment, social networks) that might cause disparate rates of contact with law enforcement, subsequently influencing exposure to use of force. Therefore, racially disparate patterns in the use of force and its severity may be spuriously related to characteristics of civilian-officer encounters that explain the use of force (e.g., Jetelina et al., 2017; Worrall et al., 2018; Knowles, Persico, and Todd, 2001;

²By “race” of the individual, we typically mean the officer’s perception of the race of the individual.

Cesario, Johnson, and Terrill, 2019). To best study the effect of one’s race on the propensity of an individual to be subjected to police use of force requires conditioning on a range of characteristics of civilians that could confound that relationship. Even still, current debates in the literature focus on the extent to which selection into contact with the police confounds our ability to draw inferences about bias from the universe of citizen-police interactions (Johnson et al., 2019; Knox and Mummolo, 2019; Knox, Lowe, and Mummolo, forthcoming).

Assuming racial bias in police shootings exists, there are at least two theoretical mechanisms, one circumstantial and the other psychological (for a brief discussion, see Ross, 2015, p. 3). The first mechanism is that racial minorities, especially Black Americans, are circumstantially associated with conditions that give rise to greater use of force by the police: They are more likely to come into contact with police because of racial profiling by officers³ or greater proximity to high-crime, highly-policed environments. The second mechanism is that police officers perceive the stakes for using force differently when they are confronting a Black civilian than when they are confronting a White civilian. In its most nefarious expression, regardless of the race of the officer, police may devalue the lives of Black civilians relative to the lives of White civilians. They might anticipate differential downstream consequences from using force against Black civilians than from using force against White civilians.

3 A Racial Bias Model of Fatal Shootings by Police

Our model builds directly from the model of Knowles, Persico, and Todd (2001) that addresses traffic stops of drivers by police. It also incorporates theory and findings about the civilian-police encounter as “a developmental process in which successive decisions and behaviors either by police officer or citizen, or both, make [a] violent outcome more or less likely” (Binder and Scharf, 1980, p. 386). Specifically, we rely on the continuum of use of force that is supposed to guide the discretion of police officers vis-a-vis use of force to propose a two-stage model of civilian-police interaction. The model seeks to capture “the transactional, or step-by-step unfolding, of police–public encounters” and the “micro process of the police-suspect encounter,” in which civilian noncompliance and resistance is often pivotal in the decisions police officer make about their discretion to use force (e.g., Terrill, 2005).

The first stage of our model is a selection stage—an interaction between a civilian and a police officer may or may not occur. At this stage, the civilian and the police officer move sequentially. First, the civilian decides whether to engage in some behavior, unlawful or lawful, that attracts the attention of the officer, either directly because the officer sees it or indirectly because another civilian reports it to the police (e.g.,

³The logic underlying racial profiling as a mechanism to explain racial disparities in use of force is potentially circular.

911). Based on the behavior, the officer might decide to investigate. The behavior could be innocuous or criminal in nature.⁴ Second, the officer decides whether to engage the civilian on the basis of the behavior.⁵ While there is obviously a wide range of behaviors an officer might engage in, we collapse this decision into a dichotomous choice between engagement or no engagement. For both the civilian and the officer, we assume the utility they receive from their behavior henceforth can be a function of any set of possible characteristics. The value of behavior for the civilian and engagement for the officer can depend on the situation, including the civilian’s own characteristics. Thus, an officer may be biased towards engaging individuals of one race, relative to another, allowing for the possibility of racially-driven selection bias in the civilian-officer interactions we observe (Knox, Lowe, and Mummolo, forthcoming). This stage of the game is important. It allows us to make empirical predictions about behavior implied by racial bias that should manifest even in the presence of strategic selection into interaction with the police. In particular, the selection stage captures, conceptually, every element of the police-civilian interaction that takes place up until the civilian and the officer reach the point of violent conflict.

In the second stage, we model a conflict subgame. We model this stage as a pair of simultaneous moves. Obviously, conflict during police-civilian encounters is the product of a series of sequential moves. We are most interested in police and civilian behavior that increases the chances that either is harmed. In our model, the myriad, slower-paced choices that civilians and officers make in the course of interaction are subsumed in the selection stage, while the simultaneous moves stage captures only those split-second choices that are made at the point of the use of force. In our view, the heightened pace of decision making, the urgency with which individuals respond to threats to their dignity and physical safety, and the uncertainty about each other (e.g., does the suspect have a gun or a bag of Skittles®) suggest this process is accurately captured by simultaneous structure. Indeed, high-profile examples of police officers shooting unarmed civilians (e.g., Amadou Diallo in the Bronx in 1999, Oscar Grant in Oakland in 2000, and Atatiana Jefferson in Fort Worth in 2019) are strongly indicative that the decision to fire a weapon is one police officers may make under conditions of uncertainty. Therefore, modeling this stage as one of simultaneous moves comports with the observed experience of OIS resulting in civilian deaths.

In our model, conflict takes the form of aggressive behavior by the civilian (actual or perceived by the officer) and the use of force by the officer, following initial interaction(s) between the civilian and officer (e.g., stopping the civilian, civilian non-compliance with verbal commands, etc.).⁶ *We assume that the likelihood*

⁴The observed behavior may be racialized, whereby some perceive the behavior as more worrisome or suspicious when its done by a civilian of a particular racial group instead of another racial group.

⁵We ignore the race of the officer and recognize an on-going scholarly debate about whether an officer’s race predicts the use of force (e.g., Alpert and Dunham, 2004; Jetelina et al., 2017).

⁶We conceive of threatening behavior by the civilian broadly—it includes posing a threat to the officer, more aggressive forms of escalating behavior, or any other kind of action that heightens the stakes of conflict between the civilian and the officer. The degree of threat by civilians against police directives, in particular, accounts for much of the variance in police use

a civilian civilian who is shot by an officer dies from the encounter increases in the extent to which the civilian acts aggressively towards the officer. (Formally, there is some smaller probability the civilian dies if s/he does not act aggressively when the officer uses force, but there is no possibility that the civilian dies if the officer does not use force.) To be clear, we use the term “aggressive” here in a very specific way—we use it to describe behavior that enhances the likelihood of death in the event of an OIS. The claim our modeling choice rests on is that a civilian should expect a greater likelihood of dying when an officer shoots if one is resisting or threatening the officer, rather than complying or deferring to the officer. This assumption is simply that in the event of a violent conflict, the police officer has an advantage, by manpower and firepower, by training and expertise.⁷ While this assumption is consequential for our analysis, we believe it is an empirically sensible one. The challenge of observing and studying what takes place in the moment during an officer-involved shooting make it hard to know precisely what causes a civilian to be more likely to die than to survive. However, some empirical studies provide reason to believe that threatening and/or resisting an officer or otherwise escalating a situation increases the likelihood that, conditional upon being in a shooting, one will die (e.g., Mohandie, Meloy, and Collins, 2009).

We assume officers and civilians alike have preferences over their behavior and physical safety. We allow the value of escalation to vary by the type of civilian. We also allow the officer’s perceptions of the cost of fatally shooting a civilian to vary by the civilian’s race. That parameter will be the basis for our main empirical “hit rate” test (e.g., Persico and Todd, 2006; Alesina and La Ferrara, 2014). In particular, we will define racial bias as a situation where officers perceive different levels of disutility associated with taking civilian lives. We will say that an officer is biased against a racial group if he thinks the relative cost of taking the life of a civilian of that group is lower than the cost of taking the life of a civilian of another group. Such a perception could be unconscious or explicitly conscious. That is, they could suffer from a simple perception that civilians of one race are more likely to be dangerous than another, or they could be consciously racist in devaluing the life of members of one race relative to another. Our decision to model this aspect of the police-civilian interaction is grounded in two sources. One, emotional reactions to threat by police officers could make them more willing to shoot a civilian (Kleider, Parrott, and King, 2010). Anxiety, in particular, makes officers more prone to shoot (Nieuwenhuys, Savelsbergh, and Oudejans, 2012). Two, the general public and police officers tend to perceive Black civilians as more threatening than non-Black civilians, regardless of situation (e.g., Welch, 2007; Correll, Park, Judd, and Wittenbrink, 2002). To the extent an officer perceives one racial group to be more threatening than another, we would expect the officer

of force during encounters with civilians (e.g., J. H. Garner, C. D. Maxwell, and Heraux, 2002).

⁷Counter-examples exist such as the 1997 North Hollywood shootout, where heavily-armed bank-robbers for a time resisted and injured officers of the Los Angeles Police Department. However, even in that example, the police fatally wounded the suspects, without suffering LAPD casualties.

to experience heightened anxiety from encountering a civilian of that race, especially if the civilian engages in non-compliance and resistance. This will shift the cost-benefit analysis associated with the risk to the officer, relative to the cost of taking the civilian’s life. We represent that shift by allowing the value of the civilian’s life to vary.

A second foundation for this representation of racial bias rests on a more dispassionate cost-benefit analysis on the part of the officer. In our model, officers incur a cost from harming a civilian, especially killing them. The first justification offered rests on the emotional or psychological costs associated with killing a potentially threatening civilian. At the same time, officers incur *anticipated* consequences from killing a civilian. These include, but are not limited to, anticipated internal affairs investigations; indictment and prosecution by a district attorney; public condemnation and stigma; professional sanctions, including loss of salary and/or benefits; or, in the rare instance, criminal sanctions, including incarceration and/or community supervision. To the extent an officer thinks he is more likely to be punished for killing a White civilian *ceteris paribus* than a Black civilian, we would say that the cost of killing the Black civilian is lower.⁸ In other words, police officers may perceive Black lives matter less than all lives, especially White lives. Of course, we also assume officers may have preferences for policing along non-racial dimensions, such as class, gender, or anything else that they might know about civilians with whom they can potentially interact.

3.1 Primitives

Players, sequence of play, and strategies. The model is played between a civilian, C , and a police officer, O . The civilian is characterized by a type, which is a pair, $\tau = \langle \kappa, \rho \rangle$. This pair includes a racial identity, $\rho \in \{B, W\}$, and a set of observable characteristics, denoted $\kappa \in \mathbb{R}$. The observable characteristics include how an individual is dressed, an individual’s demeanor, where an individual is located, with whom an individual is associated, or any other characteristic. We denote the probability density function of κ , conditional on ρ , as $g(\kappa|\rho)$. That is, the distribution of observable characteristics in the population can be different for each of the racial groups. When we turn to the empirical implications of our model, we consider a population of civilians, \mathcal{P} , characterized by the density function, $g(\cdot)$, from whom the civilian in this interaction is drawn.

The sequence of play is summarized in Figure 1 and proceeds as follows. The game begins when the civilian decides whether or not to engage in questionable or suspicious activity. Crucially, the behavior the civilian engages in need not actually be suspicious; it only need be any kind of activity that a “reasonable”

⁸Recent political developments in the U.S. might suggest that political and legal oversight is heightened when an officer kills a Black civilian. However, most attention has been focused on the *failure* to indict officers who kill Black civilians. Further, our theoretical analysis will show that the critical question is how an officer perceives the costs of killing a civilian, holding everything else constant, including the civilian’s behavior, characteristics, etc.

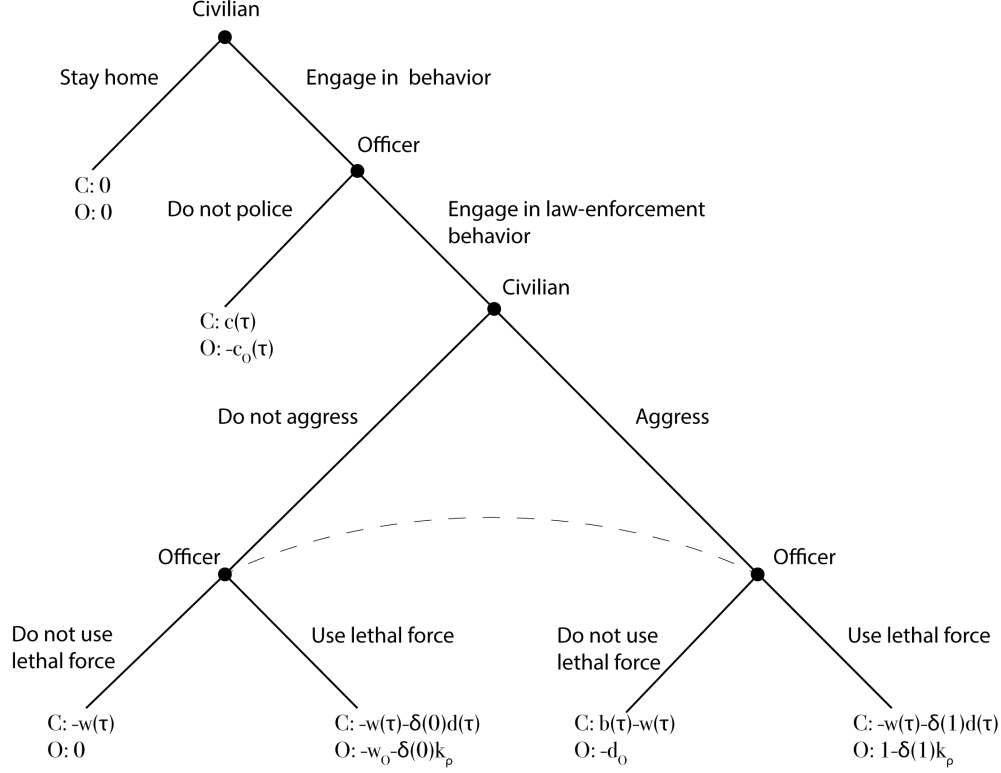


Figure 1: *Sequence of play in the model.*

officer has the ability to further investigate (e.g., “loitering” or “furtive movement”). Let $s \in \{0, 1\}$ denote that choice, where $s = 1$ indicates the choice to engage in activity potentially perceived as questionable or suspicious. If the civilian chooses $s = 0$, the game ends. However, if the civilian chooses $s = 1$, then the officer must use their discretion to decide whether to engage in law enforcement activity (e.g., stop-question-frisk). Let $l \in \{0, 1\}$ denote this choice, with $l = 1$ denoting engaging in law enforcement activity. If the officer chooses $l = 0$, the game ends; if he chooses $l = 1$, the game proceeds to the next stage, where civilian and officer actions are simultaneous. Specifically, both players must decide whether to engage in more aggressive behavior. The civilian must choose whether to aggress or not, $t \in \{0, 1\}$, where $t = 1$ denotes aggressing. The officer must choose whether to use lethal force or not, $f \in \{0, 1\}$, where $f = 1$ denotes using lethal force. If the officer chooses to use lethal force, the civilian dies with probability $\delta(t)$, where we assume $1 \geq \delta(1) > \delta(0) \geq 0$. That is, the probability the civilian dies when an officer uses lethal force is strictly greater when the civilian is aggressing than when he is not. If neither player engages in conflict (i.e., $t = 0$ and $f = 0$), then less adverse, non-fatal outcomes follow. In either event, the game ends after these choices are made, and payoffs are realized.

Let $\pi(\tau)$ denote a probability distribution over r , conditional on the civilian’s type, $\tau = \langle \kappa, \rho \rangle$, and let $\sigma(\tau)$ denote a probability distribution over f conditional on the civilian’s observable characteristics and race.

A strategy profile for the civilian is, therefore, a tuple, $\mathcal{C} = \langle s, \pi(\tau) \rangle$, and a strategy profile for the officer is a tuple, $\mathcal{O} = \langle l, \sigma(\tau) \rangle$.

Preferences and utilities. The civilian has preferences over their behavior and the outcome of their interaction with the officer. Specifically, we assume that a civilian of type τ who chooses to engage in suspicious behavior, $s = 1$, receives a payoff $c(\tau) > 0$ if the officer chooses not to engage in law enforcement activity (i.e., $l = 0$). This source of utility represents the value of engaging in whatever kind of behavior a citizen of type τ would like to engage in, without having to deal with the police. This payoff can depend on the individual's type (i.e., her race and observable characteristics). If the officer chooses to engage, though, $l = 1$, then we assume the civilian's payoff depends on whether the officer chooses to apply lethal force or not as well as whether the civilian threatens. If the officer chooses $l = 1$, then the civilian pays a cost, $-w(\tau)$, where we assume $w(\tau) > 0, \forall \tau$. This source of utility represents the cost of being subjected to law-enforcement activity and, as with the value of potentially suspicious behavior, can depend on the civilian's type. In addition to the cost of being subjected to law-enforcement activity, we assume the civilian pays a cost $-d(\tau)$ if he dies. That is, if the officer chooses to use lethal force (i.e., $f = 1$), then the civilian pays, in expectation, $-\delta(r) \cdot d(\tau)$, where we assume $d(\tau) > 0$. This source of utility represents the cost associated with losing one's life and can depend on the civilian's type—i.e., some civilians may value living more than others. To avoid considering unreasonable situations, we assume that the cost of dying is worse than the cost of being subjected to law enforcement activity for all types of civilians.

Assumption 1 (Civilians prefer not to die). $d(\tau) > w(\tau), \forall \tau$.

If the civilian aggresses, and the officer does not choose to use lethal force, then we assume the civilian receives positive utility $b(\tau) > 0$. The source of utility represents the value of engaging in resistance against an officer and can vary by type. The civilian's expected utility function is therefore given by:

$$EU_C(s, r|\tau) = \begin{cases} 0 & \text{if } s = 0 \\ c(\tau) & \text{if } s = 1 \text{ \& } l = 0 \\ -w(\tau) & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } t = 0 \text{ \& } f = 0 \\ b(\tau) - w(\tau) & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } t = 1 \text{ \& } f = 0 \\ -w(\tau) - \delta(r) \cdot d(\tau) & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } f = 1 \end{cases}$$

The officer has preferences over conducting policing work, stopping suspects and criminals, fatally wounding civilians, and his own physical well-being. Specifically, we assume the officer pays a cost $-c_O(\tau)$, where

$c_O(\tau) \in (0, 1)$, whenever the civilian chooses to engage in potentially suspicious activity (i.e., $s = 1$) and the officer does not engage in law enforcement (i.e., $l = 0$). This cost represents the cost of allowing potentially criminal activity to go overlooked or a forsaking of duty. Importantly, we allow this cost to vary by the civilian type, allowing an officer's disutility from permitting potentially criminal activity to go overlooked is a function of everything the officer can observe about the civilian. In addition, the officer pays a cost $-k_\rho$, where we assume $k_\rho \in (0, 1) \forall \rho$, whenever he fatally wounds a civilian of race ρ . Our discussion above lays out the substantive justification for this parameter, but we note its interpretation is the costs (of any kind) to taking the civilian's life. Those consequences could include emotional or psychological distress, prosecution, adverse media attention, employment consequences, or any other event that is caused by taking the civilian's life. If the officer uses force but does not fatally wound the civilian, he pays a cost $-w_O$, where $w_O > 0$. (We do not index the cost of wounding the civilian by race, but none of our results hinge on that distinction.) This cost represents expectations about the consequences of wounding a civilian (as opposed to killing the civilian), such as internal investigations, suspension from work, or any other adverse consequence that could potentially follow from using force. By contrast, the officer pays a cost, $-d_O$, where $d_O > 0$ whenever a civilian acts aggressively and he does not use lethal force, (i.e., $f = 0$). Substantively, this cost can represent injury to the officer, disutility from not stopping a criminal who is acting aggressively, or other adverse consequences. Finally, we assume the officer receives positive utility 1 from using force to stop a criminal who acts aggressively. This represents the utility of exercising authority, maintaining order, and stopping a potentially dangerous person. Therefore, the officer's expected utility function is given by:

$$EU_O(\gamma, \lambda | \tau) = \begin{cases} -c_O(\tau) & \text{if } s = 1 \text{ \& } l = 0 \\ -d_O & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } t = 1 \text{ \& } f = 0 \\ -w_O - \delta(0) \cdot k_\rho & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } t = 0 \text{ \& } f = 1 \\ 1 - \delta(1) \cdot k_\rho & \text{if } s = 1 \text{ \& } l = 1 \text{ \& } t = 1 \text{ \& } f = 1 \\ 0 & \text{otherwise} \end{cases}$$

3.2 Analysis

We characterize a mixed-strategy perfect Bayesian equilibrium. There can exist a pure strategy equilibrium if officers are never willing to use lethal force, which we rule out by assumption as implausible. For the officer to be willing to play a mixed strategy, the civilian must choose a probability distribution over her decision to aggress that makes the officer indifferent between using lethal force and not. The probability that satisfies

this requirement is:

$$\pi^*(\tau) = \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} \quad (1)$$

Notice that $\pi^*(\tau)$ is increasing in k_ρ . As an officer perceives it to be costlier to kill a civilian of race ρ , the civilian will be more likely to act aggressively. In addition, $\pi^*(\tau)$ is decreasing in $(\delta(1) - \delta(0))$. Hence, as the civilian's behavior has a larger impact on the probability of dying when the officer uses force, the equilibrium probability of a civilian threatening will decrease. Intuitively, this makes sense, for if the civilian's behavior does not matter, fatality becomes irrelevant for his calculation, and fatality is the major factor deterring him from acting aggressively. At the same time, the officer's equilibrium probability distribution over using lethal force, $\sigma^*(\tau)$, must make the civilian indifferent between choosing to aggress. That probability is given by:

$$\sigma^*(\tau) = \frac{b(\tau)}{b(\tau) + d(\delta(1) + \delta(0))} \quad (2)$$

Notice that $d(\tau) > 0, b(\tau) > 0 \implies \sigma^*(\tau) \in (0, 1)$. As long as a civilian values his life, the officer's probability of using lethal force (i.e., of choosing $f = 1$) is a proper, non-degenerate probability. If, by contrast, conditional upon being involved in a violent police situation, a civilian is indifferent to death, the officer would always use lethal force, and the civilian would die. It would not be possible to keep the civilian indifferent regarding acting aggressively (i.e., the civilian would always prefer to play $t = 1$). Substantively, we can imagine such scenarios. But we set aside this situation as theoretically uninteresting. Our definition of utility functions above assumes civilians always prefer to live than to die. Thus, in any equilibrium that reaches the conflict subgame, there exists a mixed-strategy perfect Bayesian equilibrium in which civilians probabilistically aggresses and officers probabilistically use lethal force.⁹

Proposition 1. *In any perfect Bayesian equilibrium in which the players reach the aggressive behavior subgame, the civilian and officer play mixed strategies in which a civilian of type $\tau = \langle \kappa, \rho \rangle$ chooses to threaten the police officer with probability $\pi^*(\tau)$, and the officer chooses to use lethal force with probability $\sigma^*(\tau)$.*

3.3 Empirical Implications

We now consider how racial bias by police officers could affect equilibrium behavior. We offer a simple definition of bias, guided by Knowles, Persico, and Todd, 2001. *Specifically, we say that an officer is racially biased if he perceives the cost of shooting an individual to vary by racial groups:* If an officer thinks it is less

⁹In the appendix, we show that the civilian and officer reach the conflict subgame under intuitive conditions.

costly to shoot a Black civilian than a White civilian, then we say the officer is biased against Black civilians, and vice-versa

Definition 1. *An officer is racially biased if $k_B \neq k_W$. An officer is racially unbiased if $k_B = k = k_W$.*

With this definition in hand, Proposition 1 is instructive about evidence of racial bias by police in OIS. Given Definition 1, we can identify the probability that a civilian should die, conditional upon being involved in an officer-involved shooting, when the police are not racially biased, and when they are racially biased.

Importantly, the model yields implications for how we can infer bias without having to make judgments about how to measure group traits, benefits to crime, or the distribution of traits in a group. That is, we are able to draw inferences from OIS outcomes *among those who are actually involved in a shooting*, without having data on the selection process that leads individuals into OIS events. Specifically, let $K(\rho)$ represent the set of characteristics for which an individual of race ρ would choose $s = 1$. Then, the fatality rate among people who are shot is given by

$$\mathcal{F}(\rho) = \int_{K(\rho)} \pi^*(\tau) \frac{\sigma^*(\tau) g(\kappa|\rho)}{\int_{K(\rho)} \sigma^*(z|\rho) g(z|\rho) dz} d\kappa \quad (3)$$

Notice that this fatality rate is not the fatality rate for all civilians of a given race but only for those who are shot by a police officer. Notice that by Definition 1, if an officer is not racially biased, then $k_B = k = k_W$. Given the civilian's equilibrium strategy, $\pi(\tau)^* = \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho}$, from above, then we can substitute $\frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho}$ for $\pi^*(\tau)$. Because this quantity is independent of κ , Equation (3) reduces to

$$\mathcal{F}(\rho) = \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} \quad (4)$$

Notice the only way this quantity varies with civilian race is if the officer's perceived cost of taking a civilian life varies by race. Therefore, differential fatality rates can only arise as a result of racially biased policing.

Proposition 2. *In equilibrium, different fatality rates by racial groups arise only when the officer is racially biased.*

The consequence of this result is that if police are not racially biased, then the probability an individual is killed in an officer-involved shooting, conditional on being involved in a shooting, should be independent of her race, even when taking into account all other observable characteristics that might influence her incentive to engage in noncompliance or resistance, as well as the officer's incentive to use force in the first instance. That is, Equation 3 provides the theoretical foundations for a sufficient test of racial bias in the use of lethal force in OIS. It is important to underscore that this implication of our model allows us to evaluate

evidence of racial bias, even taking into account unobservable behavioral differences across racial groups that might take place during a police-civilian confrontation. This result is parallel in logic to the way Knowles, Persico, and Todd (2001) study racial disparities in traffic stops and Alesina and La Ferrara (2014) study bias in capital sentencing. It allows us to assess evidence of racial bias without having to measure observable or behavioral characteristics of either civilians or officers. It is sufficient to evaluate variation in ultimate consequences—namely, patterns of fatality.

Implication 1. *If police officers are racially biased in favor of shooting Black civilians, then, conditional upon being involved in an officer-involved shooting, Black civilians will be less likely to die than will non-Black civilians.*

The core logic underlying this implication is that officers will be more likely to use force in less dangerous situations involving Black civilians than in similar situations involving White civilians. As a consequence, a greater proportion of OIS involving Black civilians will not lead to a fatal outcome.

4 Empirical Assessment

Our empirical assessment of the implications for racial bias in police shootings proceeds in four steps. First, we describe our method—the outcome test. Second, we describe an original dataset we built that includes all OIS (fatal and non-fatal) in nine police jurisdictions during the 2010s. Third, we focus on an evaluation of Implication 1, which predicts that racial bias among police officers will produce disparities in fatalities across racial groups. We underscore that this prediction is not intended to estimate an effect of a civilian’s race on the decision to use force; it is designed to demonstrate evidence implied by any such bias. Therefore, in the fourth step, we directly engage the issue of causal effect size. Taking our evidence as consistent with the presence of racial bias as a starting point, we calculate a lower bound for the magnitude of the effect of racial bias in the decision of an officer to shoot a civilian.

4.1 Discerning Racial Bias: The Outcome Test Method

To evaluate Implication 1, we employ an outcome or “hit rate” test, which is capable of observing disparate impact and identifying bias in decision-making (e.g., Knowles, Persico, and Todd, 2001; Persico and Todd, 2006; Alesina and La Ferrara, 2014). Mortgage lending illustrates the general logic of the approach. Mortgage lenders may care about timely repayment of loans. If we observe that non-White lenders repay mortgages on time at higher rates than Whites lenders, then that would suggest that qualified non-White applicants are being denied loans (Ayres, 2002). If the same standard were applied for lending for individuals, independent of their race, we should expect similar default rates across racial categories. However, because lenders were

willing to lend to less qualified White borrowers than to Black borrowers, the default rate would be higher for White borrowers. For policing, we may see similar systematic differences by race, in the other direction. Stops may be considered successful, for instance, if they lead to arrest, perhaps because of the discovery of contraband or the harmful behavior of drivers. Gelman, Fagan, and Kiss (2007), for example, found that 1 in 7.9 Whites police stopped were arrested, compared to 1 in 9.5 Blacks. That suggests the discretion threshold police use to decide whom to stop is lower or more indiscriminate for Black drivers than for White drivers. Our logic similarly implies that if officers have a lower threshold for deciding to shoot Black civilians than White civilians, then there will be a greater proportion of Black civilians who will choose to not threaten and, therefore, survive an officer-involved shooting.

4.2 Data on Officer-Involved Shootings

To evaluate racial disparities in fatality rates among different racial groups, we require data on every single officer-involved shooting, not just fatal shootings. Data on OIS—even just fatal ones—are notoriously difficult to acquire (Zimring, 2017). Recent efforts have begun to compile extensive data on fatal encounters between officers and civilians, though they typically rely on media reports and crowd-sourced data collection, making it difficult to assess how comprehensive and systematic the data are. Moreover, existing data typically do not include instances of OIS that do not include a fatality. Thus, we collected original data by filing public records requests with individual police departments.

We sent public records requests to the police departments and sheriff’s offices with the 50 largest jurisdictions, in terms of population, requesting records of every single instance of an officer discharging his or her weapon between 2005 and 2017. Not all agencies provided racial information about the civilians involved in these OIS. Our data therefore comprise nine jurisdictions—Charlotte; Houston; King County, WA; Los Angeles; Orlando; San Antonio; San Jose; Seattle; and Tucson—that provided comprehensive racial information in responses to our public records requests.¹⁰

The unit of analysis for each incident was the civilian/officer pair. The data the law enforcement agencies provided contained neither unique incident ID numbers nor the total number of civilian or officers involved in any incident. Therefore, to construct unique civilian/officer pairs we first made the assumption that any observations from the same city, location and date comprised the same incident. Table 1a gives an example of data we received from Tucson. Given that these two observations are both from Tucson and are on the same date of “02/07/2013” with the same location of “925 E Mill St.” we assume they are the same incident. Second, the data contained neither civilian nor officer identifying information other than race. Therefore, we

¹⁰Notably, these jurisdictions comprise eight incorporated cities and one sheriff’s office. Our results are robust to excluding the sheriff’s office.

must assume each row of data represents unique people. The incident in Table 1a would then be considered to have three total people involved: one White civilian, one Hispanic civilian and one White officer. Third, we made as many civilian/officer pairs for each incident as there were possible combinations of people. Table 1b shows how we completed the civilian/officer pairs. We now have, for example, the White civilian matched to the White officer and the Hispanic civilian matched to the White officer.¹¹

Table 1: Data Example

Obs.	Date	Location	City	Civilian Race	Officer Race
1	02/07/2013	925 E Mill St.	Tucson	White	White
2	02/07/2013	925 E Mill St.	Tucson	Hispanic	NA
(a) Example of raw data received from FOIA requests					
Obs.	Date	Location	City	Civilian Race	Officer Race
1	02/07/2013	925 E Mill St.	Tucson	White	White
2	02/07/2013	925 E Mill St.	Tucson	Hispanic	White
(b) Example of complete civilian/officer pairs					

After constructing all civilian/officer pairs we have 1292 total pairs, representing 774 unique incidents. Overall, 48% of our officer-involved shooting incidents represent fatal shootings, varying considerably by department. San Antonio had the highest rate of fatal officer-involved shooting incidents, with 13 out of 18 observations being fatal (72%). Charlotte had the lowest rate of fatalities from OIS, where 9 out of 45 observations were fatal (20%). Los Angeles had the highest number of reported OIS (663), where 58% of them were fatal. Accordingly, our data demonstrate we have considerable variation in officer-involved shooting incidents, not just by department and by time (see Figure 2) but by fatality, too.

Figure 2 shows the frequency of OIS in each of our nine jurisdictions. Because there is considerable variation in the size of the cities, there is also considerable variation in the total number of OIS. The most come from Los Angeles, the second-largest city in the country. The fewest come from San Antonio, which provided us the least comprehensive data and is also a smaller city. Therefore, we log the number of observations per month to prevent scale differences from skewing the temporal patterns and cross-jurisdiction variation. Notably, with the exception of an increase in OIS in Houston at the end of the series, we see very little within-city variation in the frequency of OIS.

What is more, the geographical distribution and concentration of OIS within cities shows intuitive but instructive patterns. Figure 3 shows the distribution of fatal and non-fatal shootings in our nine cities. Los Angeles and Houston, by far the largest cities in our dataset, experience the most OIS, whereas cities like Charlotte and Tucson experience relatively few. Additionally, it appears there is a higher fatality rate

¹¹the modal incident only has one officer-civilian pair. However, we have also estimated our model clustering observations by incident to account for within-incident correlation between observations, which does not affect the results do not change appreciably.

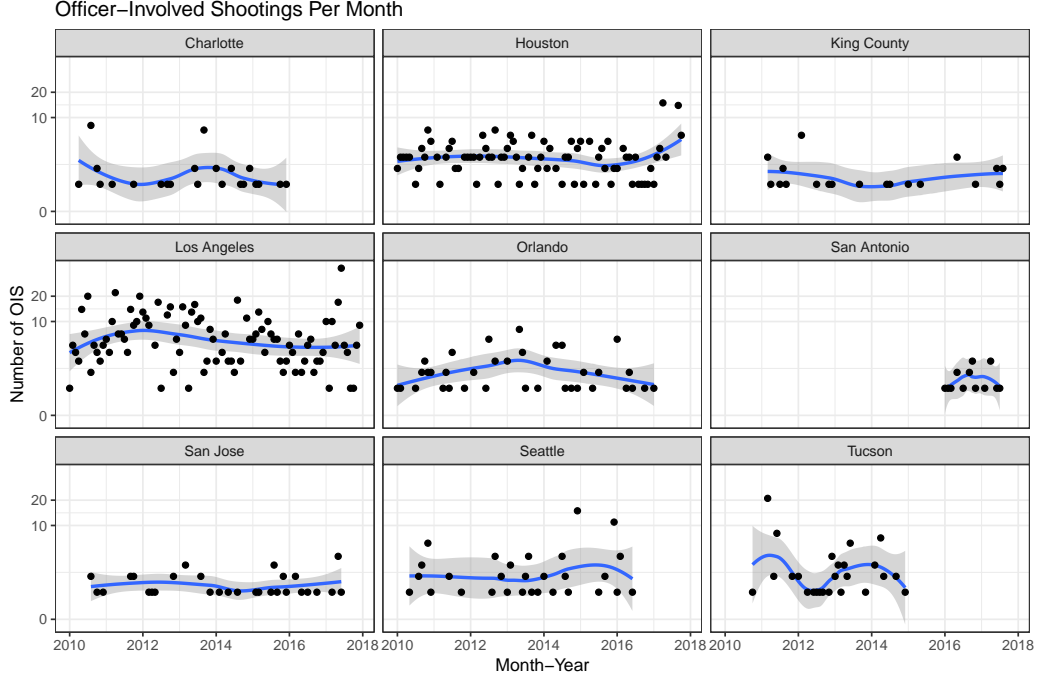


Figure 2: *Number of officer-involved shootings per month in nine cities, 2010-2017.* The figure plots the (logged) number of officer-involved shootings each month in each city.

among OIS in cities like Los Angeles and Houston, which is less of an issue in cities like San Antonio and Charlotte. Overall, Figure 3 highlights the geographical diversity in these fatal OIS, that they do not appear to systematically occur in only certain parts of certain cities, and that fatality rates vary across geographies.

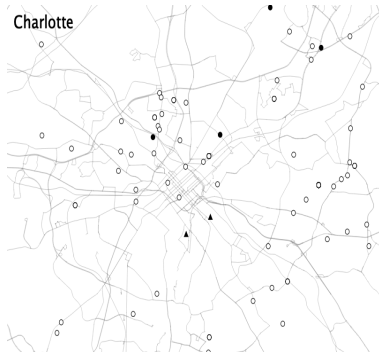
4.3 Analysis and Results

We begin our empirical analysis of Implication 1 by simply comparing the distribution of fatality across racial groups, conditional on being involved in an officer-involved shooting. Table 2 summarizes the frequencies among the observations in our data. The columns break down OIS by the race of the civilian involved, and the rows distinguish between fatal and non-fatal OIS.

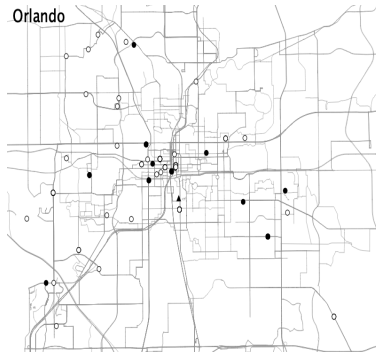
	White	Black	Hispanic	Asian
Not Fatal	120 (48%)	332 (67%)	214 (42%)	11 (27%)
Fatal	129 (52%)	165 (33%)	299 (58%)	30 (73%)

Table 2: *Summary of officer-involved shootings by race and fatality.* $\chi^2 = 77.219$, $p \leq 0.001$.

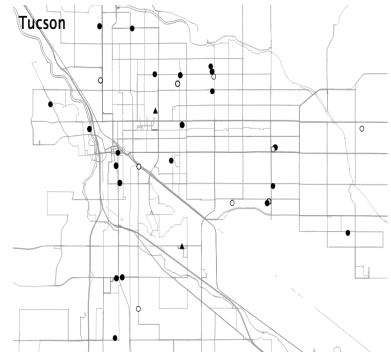
The evidence is startling, revealing considerable dependence between fatalities and the race of the civilian ($\chi^2 = 77.219$, $p \leq 0.001$). In particular, a majority of Black civilians survive OIS, whereas a majority of civilians of all other races do not. Of course, demographics and police behavior both vary across jurisdictions, and we might worry that the correlation detected in Table 2 is spurious. To speak to this we estimate a



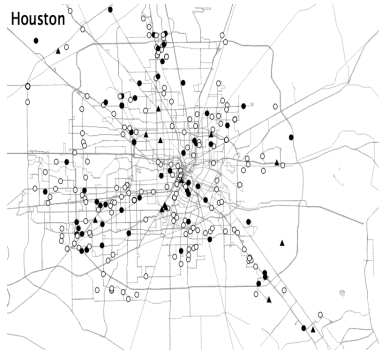
(a) Charlotte



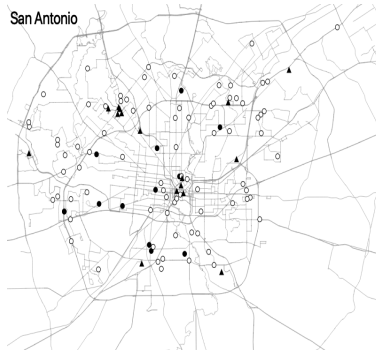
(b) Orlando



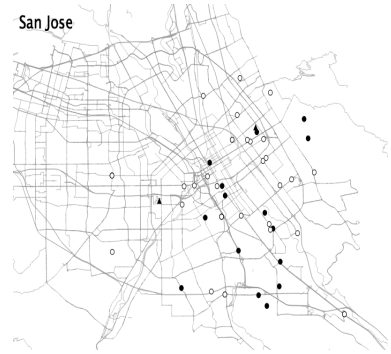
(c) Tucson



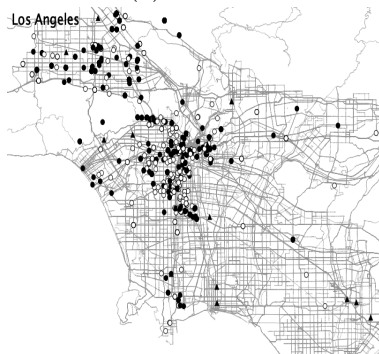
(d) Houston



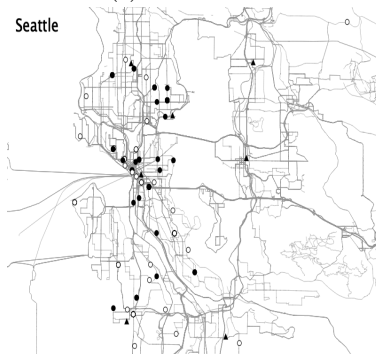
(e) San Antonio



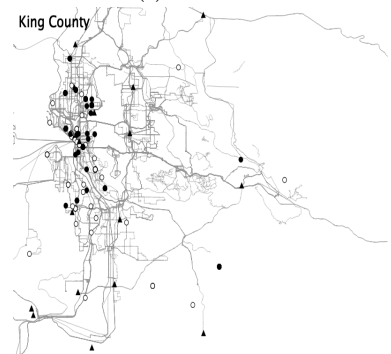
(f) San Jose



(g) Los Angeles



(h) Seattle



(i) King County

Figure 3: Locations of fatal shootings (black dots) and non-fatal shootings (white dots) in our sample of nine locations. The black triangles mark Level I Trauma Centers.

series of logistic regression specifications on all observations of OIS for which the departments we sampled provided race information. The unit of analysis is the civilian involved in an officer-involved shooting, and the outcome variable is an indicator for whether the civilian was fatally wounded. For 17 observations, the outcome was recorded as “Undetermined” or “Unknown.” We treat these observations as missing data. Our primary explanatory variable of interest is the race of the civilian involved.

We also consider specifications where we include as explanatory variables the distance from each officer-involved shooting to the nearest trauma center as well as year fixed effects (see Table 8 for the specifications with year fixed effects).¹² We also include fixed effects for the cities from which we have data, which are likely correlated with the distance to trauma center and the racial indicator. This is because trauma centers have fixed locations in cities, and demographic characteristics of populations vary across cities. Unfortunately, for 26 of our 1292 observations, the address of the officer-involved shooting was too imprecise to calculate a reliable distance measure. We consider specifications both with and without this control variable.

The main results of our analysis are reported in Table 3. The primary result appears in the top row. In each of our specifications, among those civilians shot by an officer, Black civilians are less likely to die than are White civilians. This difference is statistically significant in each specification. In our main specification, reported in the first column of results, White civilians have a predicted probability of 0.52 of dying, whereas Black civilians have a predicted probability of dying of 0.33—a 19 percentage point decrease. This relationship supports the primary empirical implication of our theoretical model of racial bias. It is consistent with the claim that police officers have a lower threshold for deciding to use lethal force against Black civilians than against White civilians. Notably, the magnitude of the relationship between being a Black civilian and the probability of dying *increases* once we include jurisdiction fixed-effects, and maintains when we include year fixed-effects. What is more, the relationship between being a Hispanic civilian and a reduced probability of dying does not emerge even after we include jurisdiction and year fixed effects. This functions as a placebo test and implies that any problematic unmeasured covariates would have to have different relationships for Black and Hispanic civilians (e.g., concerns about characteristics that affect the probability of death—such as police behavior, training, and medical attention would be largely ruled out by this analysis).

As we do not observe a depression of the relationship between being a Black civilian and the probability of survival after we include jurisdiction and time fixed effects, a spurious correlation between race and jurisdiction does not drive the observed relationship. This pattern—while not necessarily causal—is precisely

¹²Some observations lacked adequate location information to calculate the distance to the nearest trauma center, which has been shown to be a particularly important factor for the chances of survival of a gunshot wound (Crandall et al., 2013). Therefore, in the models including distance to the nearest trauma center as a control variable, we only have 1269 observations, covering 748 unique incidents.

	Model 1	Model 2	Model 3	Model 4
Black	−0.78*** (0.16)	−0.70*** (0.17)	−0.74*** (0.16)	−0.67*** (0.17)
Hispanic	0.26 (0.16)	0.07 (0.17)	0.28 (0.16)	0.10 (0.18)
Asian/AI/AN/PI	0.92* (0.37)	0.81* (0.39)	0.95* (0.38)	0.90* (0.39)
Distance			0.01 (0.01)	0.04** (0.01)
Houston		0.01 (0.41)		−0.20 (0.44)
King County		0.27 (0.55)		−0.12 (0.58)
Los Angeles		1.27** (0.40)		1.16** (0.42)
Orlando		0.59 (0.45)		0.52 (0.48)
San Antonio		1.79** (0.66)		1.87** (0.72)
San Jose		0.20 (0.51)		0.17 (0.53)
Seattle		1.25** (0.45)		1.27** (0.47)
Tucson		1.62*** (0.46)		1.65*** (0.48)
Intercept	0.08 (0.13)	−0.80* (0.40)	−0.02 (0.15)	−1.00* (0.43)
N	1292	1292	1266	1266
AIC	1718.72	1644.83	1688.72	1607.73

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3: *Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting.* Cells show logit coefficients with standard errors. Omitted category is White civilians and Charlotte. Distance is in miles.

what we expect if police are racially biased in favor of shooting Black civilians, given the logic of our model. In order to further explore the possibility that the effect is causal, we conduct a number of sensitivity analyses based on the methodology presented in Cinelli and Hazlett (2020).¹³ The sensitivity analysis considers how strong an unmeasured confounding variable would have to be in order to wipe out the effects we are finding for Black civilians. One method Cinelli and Hazlett (2020) provides to measure such strength is to benchmark any potential unmeasured confounder against measured covariates in the model. In analysis presented in the Appendix, we show that in order to render our result statistically insignificant, there would need to be an unmeasured confound that is more than three times as strong as any of the variables currently included in the model (jurisdiction fixed effects, time fixed effects, and distance to trauma center). We have not been

¹³Although this analysis is based on a linear probability model, we generally find small differences for this data between analyses based on the logit model and the linear model.

able to identify any such missing variable that would affect fatality rates for Black civilians and not Hispanic civilians.

4.4 How Big of an Effect Could Racial Bias Have on Officer-Involved Shootings?

Our analysis revealed evidence consistent with racial bias, per our definition, in the decision of police officers to use lethal force. However, we have not directly estimated a causal effect of a civilian’s race on the decision to use force. That means we still have to quantify the size of the bias, substantively. Accordingly, we estimate a lower bound on the magnitude of racial bias in OIS, relying on logic paralleling Knox, Lowe, and Mummolo (forthcoming) for identifying racial bias in police contact with civilians. The approach we adopt has three steps. First, we define the fatality rate for Black civilians that police shot, comprising two components—those shot because they were Black as opposed to White and those who would have been shot were they White or independent of their race. Second, we define the fatality rates of groups relative to each other. Third, we assume that racial bias is weakly monotonic, meaning that racial bias against Black civilians weakly increases their chances of being shot, relative to White civilians.

The magnitude of racial bias in the decision to shoot a civilian is the proportion of Black civilians shot who would not have been shot had they been White. The intuition behind this is that the observed fatality rate of Black civilians is made up of two components — Black civilians who were shot but *would not* have been shot had they been White and Black civilians who *would* have been shot had they instead been White. Our quantity of interest p , is the proportion that are in the former i.e. the proportion of Black civilians shot, who would not have been shot had they been White. By using the principle strata defining these groups as well as the weak monotonicity assumption we can derive an empirically estimable lower-bound for p .

$$p = \frac{\mathcal{F}_w - \mathcal{F}_b}{\mathcal{F}_w - \mathcal{F}_{s(b)>s(w),b}} \geq \frac{\mathcal{F}_w - \mathcal{F}_b}{\mathcal{F}_w}. \quad (5)$$

Equation (5) expresses p as a function of \mathcal{F}_w and \mathcal{F}_b , the observed fatality rates among White and Black civilians shot and $\mathcal{F}_{s(b)>s(w)}$, the fatality rate for Black civilians who would not have been shot had they been white. We do not observe $\mathcal{F}_{s(b)>s(w)}$, however, we know if there is no racial bias then $\mathcal{F}_{s(b)>s(w)}$ would be 0, as there would be no civilians shot because they were Black. Therefore, substituting 0 for $\mathcal{F}_{s(b)>s(w)}$ yields a lower bound on the true value of p . See the Appendix for formal assumptions, definitions and derivation of our quantity of interest p .

To estimate the lower bound on the proportion of Black civilians that would not have been shot had they been White, we first estimate a linear probability model. We estimate it with a subset of officer-involved shooting data containing only Black and White civilians, including our main covariate of interest, namely

race (White equal to 1, Black equal to 0), along with binary indicator variables for locality. Using this model we estimate the regression coefficient on White to be 0.17 (see the Appendix full regression specification results), the associated fatality difference between White civilians and Black civilians controlling for city fixed effects. Equation (6) defines the regression-based lower bound estimate.

$$p \geq \frac{\hat{\beta}_w}{F_b + \hat{\beta}_w} \quad (6)$$

The coefficient estimate on White is the numerator and the coefficient plus the observed mean fatality rate among Black civilians is the denominator. Thus, we estimate 33.3% is the lower bound on the proportion of Black civilians that police would not have shot had they been White.

Lower Bound Estimate	Confidence Interval around Lower Bound
0.333	[0.219,1]

Table 4: *Lower bound estimate using regression.* Lower confidence interval around the lower bound is estimated by bootstrapping. The upper interval is necessarily 1 because it could be that 100% of Black civilians would not have been shot had they been White.

Substantively, our estimate of 33.3% is considerable and given it is a lower bound, may be much higher. Our estimate implies that police would not have shot 166 Black civilians had they been White, from the 497 Black civilians in our nine localities over the years we study. Extrapolating this estimate to the larger population of the United States, however, is beyond the limits of our data. Moreover, significant intra-locality variation suggests police behavior, measured by OIS, is not uniform across the country. Additionally, comparing Hispanic civilians and Asian civilians to White civilians yielded no statistically significant differences. That is consistent with what we would expect—police officers differentially exercise discretion against Black civilians as compared to all other groups. Given the extant debate about whether the use of force by police is tainted with racial bias, these findings suggest there is a substantively significant problem. Quantifying the magnitude of its effect, though, requires richer administrative data beyond what police departments, generally, in the U.S. currently provide. Specifically, the important matter of how much police violence is attributable to racial bias requires knowing how often police fire their weapons, as well as how often they draw their weapons (e.g., Worrall et al., 2018), which is not universally known across local police departments.

5 Discussion

A significant challenge to credible inferences about the influence of racial bias in policing is that empirical observations typically need to condition on a wide range of difficult-to-measure confounds. For example, if civilian race is correlated with factors that directly affect contact with police—such as income, locality, employment rates or sectors, education level, or any possible factor—then it will be challenging to disentangle the causal effect of one’s race from the effects of those other confounding forces. However, our approach helps overcome that challenge by identifying an empirical implication of racial bias in the use of force that is conditional on contact with the police, allowing social scientists to sidestep the challenges of selection bias due to racial rates of police contact with civilians (e.g., Knox, Lowe, and Mummolo, forthcoming).

What is more, our theory, analysis, and results help make better sense of seemingly contradictory findings in the contemporary use of force literature. For example, some studies show that the probability of being Black, conditional on being shot, is not statistically different from the probability of being White, conditional on being shot (Johnson et al., 2019). In our theoretical model, however, this pattern is completely consistent with racial bias by officers in favor of shooting Black civilians. Such a pattern could emerge because Black civilians are aware of such bias and systematically avoid threatening behavior during encounters with the police that could lead to fatal OIS. Therefore, the probability of being shot, conditional on being Black, might still be higher than it is conditional on being White, even while the observed rates of being fatally wounded are the same. Similarly, our analysis can reconcile the distinction Fryer Jr. (2016) documents between lethal and non-lethal force against civilians.¹⁴ If Black civilians are aware that police officers are biased in favor of using force against them, then they should be less likely to engage in threatening behavior that would escalate a situation from a non-lethal outcome to a lethal outcome. We would expect, then, that Black civilians should be disproportionately subject to non-lethal force but not necessarily disproportionately represented in lethal encounters with police.

At the same time, while our analysis helps explain racial differences across the observed patterns in the use of force, all we can demonstrate is evidence consistent with racial bias and calculate a lower bound on the magnitude of the effect. The primary implication of our model, and the one we subject to empirical scrutiny, is a statement of an empirical regularity that is implied if civilians and officers behave as though the latter are racially biased. Lower fatality rates among Black civilians shot by the police than among White civilians shot by the police are a secondary form of evidence—a pattern implied by racial bias in the decision to shoot in the first instance. Those rates, however, do not in-and-of-themselves tell us anything about the magnitude of the effect of bias.

¹⁴Of course, Knox, Lowe, and Mummolo (forthcoming) also suggest that the analysis in Fryer Jr. (2016) is flawed due to selection bias.

However, given what we know about the existence of racial bias, we are able to calculate a lower bound on the effect size. Still, the bounds we estimate cannot tell us about the upper limit on the effect. Doing that would require we overcome the aforementioned confounding and selection challenges to inference. While not necessarily an impossible task, undertaking it remains one of the most salient limitations research on the subject faces. As we document in the appendix, our model also helps identify the kinds of assumptions or data that would be necessary to make further progress on narrowing the estimated size of racial bias in the decision to shoot a civilian.

We also note that the additional empirical implications of our analysis may lend themselves to future empirical analysis. We have not investigated Implication 2. Doing so would require objective data on observed officer interactions with civilians. In particular, we would need data on civilian behavior during all interactions with police officers, not just those involving use of force by officers. Such data are difficult to come by. However, it bears noting that there is some evidence in the extant literature that is potentially consistent with the expectation, which predicts that, if officers are racially biased against Black civilians, White civilians will be more likely to engage in escalating behavior than will Black civilians. For example, Kavanagh (1997) studies more than 1000 encounters between civilians and officers in New York’s Port Authority Bus Terminal between 1990 and 1991 and finds suggestive evidence that White civilians are more likely to resist arrest than are non-White civilians. Matrofski, Snipes, and Supina (1996) compare civilian-officer race combinations as predictors of civilian compliance with officer requests for orderly behavior. They find that, compared to White civilians interacting with White officers, White civilians interacting with minority officers are *less* likely to comply with officer instructions. At the same time, they find that minority civilians interacting with White officers are *more* likely to comply with officer instructions. They also find that minority civilians interacting with minority officers are more likely to comply, though this difference is not statistically significant. Finally, according to the FBI’s Law Enforcement Officers Killed & Assaulted data, as of July 2017, 55% of officers killed by civilians were killed by White civilians and 58% of officers assaulted by civilians were assaulted by White civilians. While far from constituting a systematic evaluation, those descriptive findings provide initial evidence to corroborate a second empirical implication of our model of racial bias.

6 Conclusion

Police-civilian encounters have special implications for the study of democratic governance and equality of citizenship. Police are perhaps the most common government official with whom civilians have contact (e.g., Jacob, 1972) and, distinct from other bureaucrats, interactions with police officers always have the

potential for violence. Consequently, the modal contact a civilian has with police relative to other government agents in the United States is one that might involve the use of physical force, including fatal and non-fatal shootings. Yet, whether justified or not, whether garnering mass and elite attention or not, whether we know enough or not about correlates and causes, police shootings (and other forms of police use of force such as pepper-spraying and canine attacks) are moments that “raise fundamental questions of governmental responsiveness and state power, and they are frequently at the heart of grievances that generate political demands and protests” (Soss and Weaver, 2016, p. 83). Police shootings, along with predatory and extractive policing (Sances and You, 2017), police “militarization” (Lawson Jr., 2019), and broader practices of policing, inclusive of surveillance, order maintenance, and arrests, coupled with choices by local prosecutors and judges (e.g., requiring bail and jailing arrestees for low-level offenses), invite political scientists to ask “questions about police authority, state projects of social control, and daily encounters with local governance” (Soss and Weaver, 2017, p. 568). They also invite questions about the influence of bias, especially racial bias.

Racial bias on the part of government officials has the distinct potential to undermine the legitimacy of the state and civilian cooperation and engagement with government. To the extent, then, that police officers engage in racially biased use of force, that behavior has potentially profound consequences for the maintenance of a well-functioning democratic order. In light of these observations, recent analyses of racial disparities in the use of force by police officers have set out to address whether and how much racial bias influences policing in the United States. The implications of the findings are far-reaching.

Our results raise concern about racial bias in the use of force by police. They also highlight the need for more research and more comprehensive data about OIS, including, among other things, officer attributes and situational and contextual factors. For example, to understand the mechanisms by which racial bias affects civilian and police behavior, scholars need to study all civilian interactions with police, not just those encounters ending in fatalities, or even just the encounters where the use of force occurred. Of course, as others have pointed out (e.g., Knox, Lowe, and Mummolo, forthcoming) and as our model considers, there is potentially racial bias in the initial selection of civilians into contact with police. To the extent racial bias systematically affects not just how police interact with civilians but with which civilians they interact, our analysis underscores the extent to which training, recruiting, and monitoring of police officers have implications beyond public and officer safety.

Although our empirical study provides evidence consistent with racial bias in the use of force and a lower bound on the magnitude of racial bias in the decision to shoot, more research is necessary to assess the magnitude of the effect. We also need more research on racial bias in policing to assess the efficacy of policies designed to minimize racial disparities in policing, as well as to determine the underlying mechanisms that produce such racial bias. While normatively we might believe that, independent of its cause, racial disparities

are problematic, what to do about them depends on identifying the root cause. In particular, whether racial disparities are a result of circumstantial factors or systematic bias by police officers affects what kinds of remedies are desirable and the implications of the disparities for the legitimacy and integrity of the police as a key law enforcement institution.

But better research will require richer administrative data on police practices, ranging across the use of force continuum, and outcomes (i.e., weapons drawn, weapons fired, and lethal and non-lethal consequences). The current nature and contents of use of force and consequences record-keeping by many police departments, however, presents serious challenges to improving research and establishing consensus in weighting across the varied factors associated with officer-involved shootings, complicated further by the decentralization of law enforcement and its discretion across localities in the United States. Nonetheless, we expect that as police departments, elected officials, and institutions of civilian oversight of police departments become more interested in research about policing practices and outcomes, more anticipatory of scholarly needs, more transparent about and willing to share their data with scholars and others through digitization and open-access, along with replication and extension of academic studies, causal research on police behavior, from the spectacular to the mundane, and better policymaking to improve policing (and police legitimacy) will flourish.

References

- Alesina, Alberto and Eliana La Ferrara (2014). “A Test of Racial Bias in Capital Sentencing”. In: *American Economic Review* 104.11, pp. 3397–3433.
- Alpert, Geoffrey P. and Roger G. Dunham (2004). *Understanding Police Use of Force: Officers, Suspects, and Reciprocity*. Cambridge University Press.
- Ayres, Ian (2002). “Outcome Tests of Racial Disparities in Police Practices”. In: *Justice Research and Policy* 4.1-2, pp. 131–142.
- Bateson, Regina (2012). “Crime victimization and political participation”. In: *American Political Science Review* 106.3, pp. 570–587.
- Baumgartner, Frank R., Derek A. Epp, and Kelsey Shoub (2018). *Suspect Citizens: What 20 Million Traffic Stops Tell Us About Policing and Race*. Cambridge University Press.
- Binder, Arnold and Peter Scharf (1980). “The Violent Police-Citizen Encounter”. In: *The ANNALS of the American Academy of Political and Social Science* 452.1, pp. 111–121.
- Brown, Robert A (2019). “Policing in American History”. In: *Du Bois Review: Social Science Research on Race* 16.1, pp. 189–195.

- Buehler, James W. (2017). "Racial/Ethnic Disparities in the Use of Lethal Force by US Police, 2010–2014". In: *American Journal of Public Health* 107.2, pp. 295–297.
- Canner, Glenn B., Stuart A. Gabriel, and J. Michael Woolley (1991). "Race, Default Risk and Mortgage Lending: A Study of the FHA and Conventional Loan Markets". In: *Southern Economic Journal*, pp. 249–262.
- Cesario, Joseph, David J. Johnson, and William Terrill (2019). "Is There Evidence of Racial Disparity in Police Use of Deadly Force? Analyses of Officer-Involved Fatal Shootings in 2015–2016". In: *Social Psychological and Personality Science* 10.5, pp. 586–595.
- Cinelli, Carlos and Chad Hazlett (2020). "Making sense of sensitivity: Extending omitted variable bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67.
- Correll, Joshua, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink (2002). "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals". In: *Journal of Personality and Social Psychology* 83.6, p. 1314.
- Correll, Joshua, Bernadette Park, Charles M. Judd, Bernd Wittenbrink, et al. (2007). "Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot". In: *Journal of Personality and Social Psychology* 92.6, p. 1006.
- Crandall, Marie et al. (2013). "Trauma Deserts: Distance from a Trauma Center, Transport Times, and Mortality from Gunshot Wounds in Chicago". In: *American Journal of Public Health* 103.6, pp. 1103–1109.
- Eberhardt, Jennifer L. et al. (2004). "Seeing Black: Race, Crime, and Visual Processing". In: *Journal of Personality and Social Psychology* 87.6, p. 876.
- Engel, Robin Shepard and Jennifer M. Calnon (2004). "Examining the Influence of Drivers' Characteristics During Traffic Stops with Police: Results from a National Survey". In: *Justice Quarterly* 21.1, pp. 49–90.
- Fryer Jr., Roland G. (2016). *An Empirical Analysis of Racial Differences in Police Use of Force*. Tech. rep. National Bureau of Economic Research.
- Garner, Joel H., Christopher D. Maxwell, and Cedrick G. Heraux (2002). "Characteristics Associated with the Prevalence and Severity of Force Used by the Police". In: *Justice Quarterly* 19.4, pp. 705–746.
- Garner, Joel H., Thomas Schade, et al. (1995). "Measuring the Continuum of Force Used by and Against the Police". In: *Criminal Justice Review* 20.2, pp. 146–168.
- Garner, Joel and Christopher Maxwell (1999). *Measuring the Amount of Force Used by and Against the Police in Six Jurisdictions*. Tech. rep. Bureau of Justice Statistics and National Institute of Justice.

- Gelman, Andrew, Jeffrey Fagan, and Alex Kiss (2007). "An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias". In: *Journal of the American Statistical Association* 102.479, pp. 813–823.
- Goel, Sharad, Justin M. Rao, and Ravi Shroff (2016). "Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy". In: *The Annals of Applied Statistics* 10.1, pp. 365–394.
- Goff, Philip Atiba et al. (2016). *The Science of Justice: Race, Arrests, and Police Use of Force*. Center for Policing Equity.
- Jacob, Herbert (1972). "Contact with Government Agencies: A Preliminary Analysis of the Distribution of Government Services". In: *Midwest Journal of Political Science*, pp. 123–146.
- James, Lois, Bryan Vila, and Kenn Daratha (2013). "Results from Experimental Trials Testing Participant Responses to White, Hispanic and Black Suspects in High-Fidelity Deadly Force Judgment and Decision-Making Simulations". In: *Journal of Experimental Criminology* 9.2, pp. 189–212.
- Jetelina, Katelyn K et al. (2017). "Dissecting the complexities of the relationship between Police Officer–civilian race/Ethnicity dyads and less-than-Lethal Use of Force". In: *American Journal of Public Health* 107.7, pp. 1164–1170.
- Johnson, David J. et al. (2019). "Officer Characteristics and Racial Disparities in Fatal Officer-Involved Shootings". In: *Proceedings of the National Academy of Sciences*, pp. 15877–15882.
- Kahn, Kimberly Barsamian et al. (2017). "How suspect race affects police use of force in an interaction over time." In: *Law and human behavior* 41.2, p. 117.
- Kavanagh, John (1997). "The Occurrence of Resisting Arrest in Arrest Encounters: A Study of Police-Citizen Violence". In: *Criminal Justice Review* 22.1, pp. 16–33.
- Kleider, Heather M., Dominic J. Parrott, and Tricia Z. King (2010). "Shooting Behaviour: How Working Memory and Negative Emotionality Influence Police Officer Shoot Decisions". In: *Applied Cognitive Psychology* 24.5, pp. 707–717.
- Knowles, John, Nicola Persico, and Petra Todd (2001). "Racial Bias in Motor Vehicle Searches: Theory and Evidence". In: *Journal of Political Economy* 109.1, pp. 203–229.
- Knox, Dean, Will Lowe, and Jonathan Mummolo (forthcoming). "Administrative Records Mask Racially Biased Policing". In: *American Political Science Review*.
- Knox, Dean and Jonathan Mummolo (2019). "Making Inferences about Racial Disparities in Police Violence". Princeton University Working Paper.
- Lawson Jr., Edward (2019). "TRENDS: Police Militarization and the Use of Lethal Force". In: *Political Research Quarterly* 72.1, pp. 177–189.

- Matrofski, Stephen D., Jeffrey B. Snipes, and Anne E. Supina (1996). "Compliance on Demand: The Public's Response to Specific Police Requests". In: *Journal of Research in Crime and Delinquency* 33.3, pp. 269–305.
- Mekawi, Yara and Konrad Bresin (2015). "Is the evidence from racial bias shooting task studies a smoking gun? Results from a meta-analysis". In: *Journal of Experimental Social Psychology* 61, pp. 120–130.
- Mohandie, Kris, J. Reid Meloy, and Peter I Collins (2009). "Suicide by Cop Among Officer-Involved Shooting Cases". In: *Journal of Forensic Sciences* 54.2, pp. 456–462.
- Mummolo, Jonathan (2018). "Modern Police Tactics, Police-Citizen Interactions, and the Prospects for Reform". In: *The Journal of Politics* 80.1, pp. 1–15.
- Munnell, Alicia H. et al. (1992). "Mortgage Lending in Boston: Interpreting the Data". In: *Federal Reserve Bank of Boston, Oct. Working Paper*, pp. 92–7.
- Nieuwenhuys, Arne, Geert J.P. Savelsbergh, and Raoul R.D. Oudejans (2012). "Shoot or Don't Shoot? Why Police Officers Are More Inclined to Shoot When They Are Anxious". In: *Emotion* 12.4, p. 827.
- Persico, Nicola and Petra Todd (2006). "Generalising the Hit Rates Test for Racial Bias in Law Enforcement, with an Application to Vehicle Searches in Wichita". In: *The Economic Journal* 116.515, F351–F367.
- Pierson, Emma et al. (2017). "A Large-Scale Analysis of Racial Disparities in Police Stops across the United States". Stanford University Working Paper.
- 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". In: *PLoS One* 10.11, e0141854.
- Sances, Michael W. and Hye Young You (2017). "Who Pays for Government? Descriptive Representation and Exploitative Revenue Sources". In: *The Journal of Politics* 79.3, pp. 1090–1094.
- Schuck, Amie M. (2004). "The Masking of Racial and Ethnic Disparity in Police Use of Physical Force: The Effects of Gender and Custody Status". In: *Journal of Criminal Justice* 32.6, pp. 557–564.
- Sikora, Andrew G. and Michael Mulvihill (2002). "Trends in Mortality Due to Legal Intervention in the United States, 1979 Through 1997". In: *American Journal of Public Health* 92.5, pp. 841–843.
- Soss, Joe and Vesla Weaver (2016). "Learning from Ferguson". In: *The Double Bind: The Politics of Racial & Class Inequalities in the Americas*. Report of the APSA Task Force on Racial & Class Inequalities in the Americas.
- (2017). "Police Are Our Government: Politics, Political Science, and the Policing of Race-Class Subjugated Communities". In: *Annual Review of Political Science* 20, pp. 565–591.
- Terrill, William (2005). "Police Use of Force: A Transactional Approach". In: *Justice Quarterly* 22.1, pp. 107–138.
- (2011). "Police Coercion: Application of the Force Continuum". Ph.D. Dissertation.

- Voigt, Rob et al. (2017). “Language from Police Body Camera Footage Shows Racial Disparities in Officer Respect”. In: *Proceedings of the National Academy of Sciences* 114.25, pp. 6521–6526.
- Weaver, Vesla and Amy Lerman (2010). “Political Consequences of the Carceral State”. In: *American Political Science Review* 104.4, pp. 817–833.
- Welch, Kelly (2007). “Black Criminal Stereotypes and Racial Profiling”. In: *Journal of Contemporary Criminal Justice* 23.3, pp. 276–288.
- White, Ariel (2019). “Misdemeanor Disenfranchisement? The demobilizing effects of brief jail spells on potential voters”. In: *American Political Science Review* 113.2, pp. 311–324.
- Worden, Robert E. (2015). “The ‘Causes’ of Police Brutality: Theory and Evidence on Police Use of Force”. In: *Criminal Justice Theory: Explaining The Nature and Behavior of Criminal Justice* 2, pp. 149–204.
- Worrall, John L. et al. (2018). “Exploring Bias in Police Shooting Decisions With Real Shoot/Don’t Shoot Cases”. In: *Crime & Delinquency*, pp. 1171–1192.
- Zimring, Franklin E. (2017). *When Police Kill*. Harvard University Press.

A Supplemental Theoretical Results

In any equilibrium in which officers choose to engage in law-enforcement activity—i.e., any equilibrium that reaches the aggressive behavior subgame—there can exist one pure strategy equilibrium to the game, but it requires a particularly strong condition. Specifically, there can exist a pure strategy equilibrium, where the aggressive behavior subgame involves the civilian always choosing to threaten ($t = 1$) and the officer choosing never to use lethal force ($f = 0$) if the officer views the cost of killing a civilian is much larger than the cost of losing his own life—formally, if $\delta(1)k_\rho - d_O > 1$. In other words, the officer must regard the differential between the value of his own life and the life of the civilian as being greater than the value of stopping crime. That is, the officer would never be willing to use force to stop a violent criminal.

Lemma 1. *Any pure strategy equilibrium to the aggressive behavior subgame involves the civilian always threatening ($t = 1$) and the officer never using lethal force ($f = 0$). This equilibrium can only hold for $\delta(1)k_\rho - d_O > 1$; that is, when the officer regards the difference between value of the civilian’s life and his own to be greater than the value of stopping a violent criminal.*

Lemma 1 shows that any pure strategy equilibrium is substantively uninteresting. It can only occur under conditions where an officer is completely unwilling to use lethal force. In addition, pure strategy equilibria are not substantively interesting insofar as we observe variation in civilian and police officer behavior, conditional on both observable characteristics and racial categories. In the observed world, officers and civilians appear to be playing mixed strategies. We, therefore, focus the remainder of our analysis on characterizing a mixed strategy equilibrium. We assume, then, that the officer is always willing to use lethal force, if necessary (Alpert and Dunham, 2004).

Assumption 2 (Officers willing to use force). *The officer is always weakly willing to use lethal force, $\delta(1)k_\rho - d_O \leq 1, \forall \rho$*

In addition to our theoretical model’s implications for observable implications of racial bias, it also yields insights about the extent to which racial bias might affect what we can learn from studying police-civilian contact altogether. Starting with the officer’s decision to engage in law-enforcement activity, the model reveals a number of factors that are important. First, as we described, in equilibrium we will only observe interactions between individuals whose behavior is sufficiently costly to ignore and the police. Specifically, it must be the case that $c_O(\tau)$ —the cost of overlooking potentially criminal activity behavior—is sufficiently large in order to observe law-enforcement activity. While we allow this parameter to vary by the civilian type, we do not assume that observable characteristics and race are orthogonal. That means that evidence of racial bias from the rate of police engagement with a population must take the form of racial disparities *conditional on all observable characteristics*. Moreover, because once an officer decides to engage in law-enforcement

activity, the probability the officer escalates (i.e., uses force) will also be driven by civilian characteristics. (Recall, the officer's strategy must keep the civilian indifferent between threatening and not threatening.) Therefore, absolute rates of contact between officers and the use of force across racial categories cannot in and of themselves demonstrate racial bias by the police (see also, Knox, Lowe, and Mummolo, forthcoming).

To assess the effect of racial bias on observed police-civilian interactions, we therefore need to evaluate how changing the value of k_ρ affects each stage of the game. we can re-arrange Condition (8) as follows:

$$\pi(\tau)^* \leq \frac{w(\tau) + \sigma^* d\delta(0)}{b(\tau)(1 - \sigma^*) - \sigma^* d(\delta(1) - \delta(0))}$$

Because $\frac{\partial \pi(\tau)^*}{\partial k_{rho}} > 0$, as the cost of killing a civilian of race ρ increases, the civilian of that race, holding constant his observable characteristics, κ , is more willing to play $s = 1$. And, as we saw, Condition (7) does not depend on k_ρ .

Therefore, as police become increasingly racially biased against civilians of race ρ , we should see the pool of individuals interacting with the police shift. In particular, if $k_W > k_B$ —that is, if the police are racially biased against civilians of race B , then, conditional on observable characteristics, we should see more civilians of race W being subjected to law-enforcement and ultimately killed by police. The intuition here is that individuals of race B will censor their behavior in anticipation of a lower threshold by police for using force against them.

Proposition 3. *Racial bias by police induces selection bias in the observed population of civilian-police interactions. All else equal, civilians of the race against whom police are biased censor their behavior and are less likely to engage in behavior that could trigger law-enforcement activity than are civilians of another race.*

Of course, Proposition 3 is an all-else-equal statement. It must be the case not only that κ is held constant, but so, too, are the other parameters, especially $w(\tau)$, $b(\tau)$, and $c(\tau)$. This means that the costs of being stopped by the police, the benefit of resisting police, and the benefits of engaging in potentially suspicious activity must be held constant. There is good reason to believe that these quantities are correlated with race, even holding constant observable characteristics, because of the unique social, political, and economic experiences that people of different races have in the US. The consequence is that even attempting to control for every observable characteristics of civilians who could potentially be subjected to law-enforcement activity will not alleviate selection bias (cf, Knox, Lowe, and Mummolo, forthcoming).

Now we turn to the first stage of the game to assess the conditions under which the players reach the subgame where they decide whether to threaten and use lethal force. For the officer to choose $l = 1$, it must

be the case that the equilibrium expected utility from reaching the aggressive behavior stage is better than the cost of letting a suspicious civilian or possible criminal go undeterred. Abusing notation, this condition is given by:

$$c_O(\tau) \geq \frac{d_O b(\tau)}{b(\tau) + d(\delta(1) + \delta(0))} \quad (7)$$

This condition can be met under a variety of conditions. Most simply, the officer will be willing to engage in law enforcement activity (i.e., play $l = 1$) for any arbitrarily large value of $c_O(\tau)$ —that is, if the officer finds it too costly to ignore potentially suspicious activity by a civilian of type τ . Later, we consider the empirical implications of this result, but we note that if officers are more suspicious of individuals of a particular race or with a particular set of observable characteristics, we might expect disproportionate engagement by the police with those types of individuals, even for the same type of behavior. This pattern could give rise to observed disparities in the racial makeup of individuals stopped by police and selection bias in the observed set of civilian-officer interactions (see, for example, Knox, Lowe, and Mummolo, forthcoming).

Finally, then, we show that the civilian can have a positive expected utility from the subgame, which is sufficient to induce her to be willing to enter into the confrontation in the first instance. This condition is met whenever the following is satisfied:

$$\sigma^* [\pi^*(\tau)(\delta(1) - \delta(0)) + \delta(0)] \leq \frac{\pi^*(\tau) b(\tau) (1 - \sigma^*) - w(\tau)}{d} \quad (8)$$

or, alternatively, whenever the officer is unwilling to engage in law-enforcement activity—i.e., when Condition (7) is not satisfied.

While Condition (8) is algebraically messy, it has an intuitive interpretation. It can be satisfied when either (a) $\delta(1) - \delta(0)$, (b) d , or (c) $w(\tau)$ is small enough, relative to other parameters. Substantively, that means a civilian is willing to engage in potentially suspicious behavior whenever Recall, the behavior need not actually be suspicious, and we may think of this as a minimal condition. If a civilian compares essentially remaining cloistered to living a free life and finds the value of engaging in his life's activities to be sufficiently valuable relative to the inconvenience of potentially being stopped by the police, along with the subsequent potential outcomes, then he will be willing to engage in said activity.

Proposition 4. *There exists a unique perfect Bayesian equilibrium in which a civilian of type $\tau = \langle \kappa, \rho \rangle$ initiates conflict for sufficiently low $w(\tau)$ or large d_C , the officer chooses to engage the civilian for sufficiently low values to either the officer or the civilian of life, and the civilian and officer probabilistically threaten and use force, with probabilities $\pi(\tau)^*$ and $\sigma(\tau)^*$, respectively.*

Proposition 4 provides a number of insights into the nature of officer-involved shootings. First, it suggests

there can be disparities in who is involved in officer-involved shootings that may or may not be driven by racial bias. A sufficient condition for a racial disparity is that the distribution of characteristics that make the value of life lower for civilians is higher for individuals of one racial group than another. Similarly, it might be that the distribution of characteristics that make the value of crime high is larger for one group than the other. In other words, Proposition 4 reveals that the source of racial disparities in the prevalence of officer-involved shootings (fatal and non-fatal) cannot be ascertained without identifying the distribution of characteristics in the population associated with how individuals value life and crime. However, as we show in the next section, there are implications that follow from this model that allow us to assess racial bias without having to measure such concepts.

At the same time, Proposition 2 reveals a related, but distinct, empirical implication. Note that civilian behavior is designed to maintain indifference by the officer. Therefore, civilians of a race for whom officers are biased in favor of using lethal force should be less likely to threaten officers, *ceteris paribus*. That is, for the same reason we expect Black civilians to be more likely to survive a police officer's use of force, we expect, conditional on being subjected to law-enforcement activity, White civilians will be more likely to engage in threatening behavior, such as resisting arrest, disobeying officer commands, or behaving belligerently. There is no definitive evidence, however, from empirical studies of suspect resistance to support the expectation. While some studies find that Whites are more likely to escalate their behavior during encounters with the police, some studies suggest civilian of color are more likely to escalate towards harm.¹⁵

To see this, note that

$$\frac{\partial \pi^*(\tau)}{\partial k_\rho} = \frac{\delta(0)}{1 + d_O + w_O - (\delta(1) - \delta(0))k_\rho} + \frac{(\delta(1) - \delta(0))(w_O + \delta(0)k_\rho)}{(1 + d_O + w_O - (\delta(1) - \delta(0)))^2} > 0$$

That is, as the cost of taking a civilian's life increases, so too does the probability that a civilian of that race threatens the officer. Therefore, given Definition 1, if an officer is racially biased towards killing a civilian of race ρ (i.e., $k_\rho < k_{-\rho}$), then that civilian should be less likely to engage in behaviors during the encounter that threaten the officer or elevate the use of force by the officer. Conversely, if an officer is racially biased against killing civilian of some race, then that civilian should be more likely to engage in behaviors during the encounter that threaten the officer or elevate the use of force by the officer.

¹⁵Empirical studies directly assessing whether race of civilian is associated with civilian non-compliance and resistance during encounters with police are few. Their conclusions, derived from a variety of sources, including post-encounter narratives of police, use of force case reports, and surveys of victimized and non-victimized police officers, are mixed. Some studies observe that Whites may be quicker than non-Whites to display resistance during encounters with police (Kahn et al., 2017). Others observe no differences by race in civilian resistance to the police (e.g., Bierie, Detar, and Craun, 2016). The remainder claim race-based differences, with Blacks being more likely to be (or be perceived) as resistant during police-civilian encounters (Belvedere, John L Worrall, and Tibbetts, 2005; Bierie, 2017). In sum, there is no scholarly consensus about race as an explanation for civilian resistance during police encounters.

Implication 2. *If police officers are racially biased against shooting White civilians, then, conditional upon being subjected to law-enforcement activity, White civilians will be more likely to engage in threatening behavior, such as resisting arrest, disobeying officer commands, or behaving belligerently than will non-White civilians.*

B Proofs of Formal Results

Proof of Lemma 1. The proof proceeds in two steps. First, we show that no pure strategy pair other than $\langle t = 1, f = 0 \rangle$ can be an equilibrium. Second, we show that $\langle t = 1, f = 0 \rangle$ can be an equilibrium on for $k_\rho - d_O > 1$.

Consider the strategy pair $\langle t = 1, f = 1 \rangle$. The civilian receives the payoff $-w(\tau) - d(\tau)$. By deviating to $t = 0$, the civilian receives $-w(\tau)$, which is strictly greater and so would not play $\rho = r$ in equilibrium. Next, consider the strategy pair $\langle r = 0, f = 1 \rangle$. The officer receives a payoff $-w_O - c_O$. By deviating to $f = 0$, the officer receives 0, which is strictly greater, and so this strategy pair cannot be an equilibrium. Now, consider the pair $\langle t = 0, f = 0 \rangle$. The civilian receives $-w(\tau)$. By deviating to $t = 1$, she receives $b(\tau) - w(\tau)$, which is strictly greater, and therefore the strategy pair cannot be an equilibrium.

Next, consider the strategy pair $\langle t = 1, f = 0 \rangle$. The officer receives utility $-d_O$, and the civilian receives utility $b(\tau) - w(\tau)$. By deviating to $t = 0$, the civilian receives $-w(\tau)$ and so has no incentive to deviate. By deviating to $f = 0$, the officer receives $1 - k_\rho$. Thus, he has an incentive to deviate if only if $k_\rho - d_O > 1$. \square

Proof of Proposition 1. In order for the civilian and the officer to play mixed strategies in a perfect Bayesian equilibrium, each player's strategy must make the other indifferent between the elements of her choice set. This means the civilian's probability distribution over t must satisfy:

$$\begin{aligned} EU_O(f = 1|\pi^*, \tau) &= EU_O(f = 0|\pi^*, \tau) \\ \pi^*(\tau)(1 - \delta(1)k_\rho) - (1 - \pi^*(\tau))(w_O + \delta(0)k_\rho) &= -\pi^*(\tau)d_O \\ \pi^*(\tau) &= \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} \end{aligned}$$

Similarly, the officer's equilibrium probability distribution over f must make the civilian indifferent over the elements of her choice set, t . That is, $\sigma^*(\tau)$ must solve

$$\begin{aligned} EU_i(t = 1|\sigma, \tau) &= EU_i(t = 0|\sigma, \tau) \\ -\sigma^*(\tau)(w(\tau) + \delta(1)d(\tau)) + (1 - \sigma^*(\tau))(b(\tau) - w(\tau)) &= -\sigma^*(\tau)(w(\tau) + \delta(0)d(\tau)) - (1 - \sigma^*(\tau))w(\tau) \\ \sigma^*(\tau) &= \frac{b(\tau)}{b(\tau) + d(\delta(1) + \delta(0))} \end{aligned}$$

Assumption 2 $\implies \pi^*(\tau) \in (0, 1)$, and Assumption 1 $\implies \sigma^*(\tau) \in (0, 1)$. Therefore, for all parameter values, the players can make each other indifferent and can therefore play mixed strategies in equilibrium. \square

Proof of Proposition 4. Proposition 1 demonstrates that in the subgame where the officer has decided to engage in law-enforcement activity (i.e., $l = 1$), there exists a mixed strategy perfect Bayesian equilibrium. To complete the proof, we must show that the subgame can be reached in equilibrium and that the mixed strategies characterized by Proposition 1 constitute the unique equilibrium.

In order to reach the subgame, the officer must be willing to engage in law-enforcement activity, and the civilian must be willing to engage in potentially suspicious behavior. A sufficient condition for the civilian to play $s = 0$ is $EU_i[s = 1, \pi|\tau] \geq EU_i[s = 0, \pi|\tau]$ and a sufficient condition for the officer to play $l = 1$ is $EU_O[l = 1, \sigma|\tau, s] \geq EU_O[l = 0, \sigma|\tau, s]$. Notice that if $s = 0$, the game ends. Therefore, we it is sufficient to show

$$\begin{aligned} EU_O[l = 1|\tau, s = 1] &\geq EU_O[l = 0|\tau, s = 1] \\ -\sigma^*(\pi^*(1 + \delta(1)k_\rho) + (1 - \pi^*)(w_O + \delta(0)k_\rho)) - (1 - \sigma^*)d_O &\geq -c_O(\tau) \\ \frac{d_O b(\tau)}{b(\tau) + d(\delta(1) + \delta(0))} &\leq c_O(\tau) \end{aligned}$$

which can be true for an arbitrarily large value of $c_O(\tau)$. Finally, we must show that, given these constraints, the civilian is willing to engage in potentially suspicious behavior. Formally, it must be the case that

$$\begin{aligned} EU_i[s = 1|\tau] &\geq EU_i[s = 0|\tau] \\ -\sigma^*[\pi^*(w(\tau) + \delta(1)d(\tau)) - (1 - \pi^*)(w(\tau) + \delta(0)d)] + (1 - \sigma^*)[\pi^*(b(\tau) - w(\tau)) - (1 - \pi^*)w(\tau)] &\geq 0 \\ w(\tau) &\leq d \left[\frac{\pi^*(\tau)b(\tau)(1 - \sigma^*)}{d} - \sigma^*[\pi^*(\tau)(\delta(1) - \delta(0)) + \delta(0)] \right] \end{aligned}$$

which can be true for an arbitrarily small value of $w(\tau)$. Now to see that the equilibrium is unique, notice that Assumption 2 and Lemma 1 imply the mixed strategies $\pi^*(\tau)$ and $\sigma^*(\tau)$ are the unique equilibrium strategies in the aggressive behavior subgame. Further, notice the earlier stage of the game involves perfect and complete information, and the players cannot be indifferent between their strategies choices. Therefore, the pure strategies given by Conditions (7) and (8) characterize the unique equilibrium. \square

Proof of Proposition 2. Fatality rates are the proportion of civilians who die among those for whom O chooses to play $f = 1$; therefore, it is directly proportional to $\pi^*(\tau)$. In order for there to be different fatality proportions among racial groups, $\pi^*(\tau)$ must vary by ρ . The only parameter in $\pi^*(\tau)$ that is a function of ρ is k_ρ . Notice that by Definition 1, if an officer is not racially biased, then $k_B = k = k_W$. Given

the equilibrium probability of choosing to threaten is $\pi^*(\tau) = \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho}$, from above, then we can substitute $\frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho}$ for $\pi^*(\tau)$ in Equation (3) and re-arrange as follows:

$$\begin{aligned}\mathcal{F}(\rho) &= \int_{K(\rho)} \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} \frac{\sigma^*(\tau)g(\kappa|\rho)}{\int_{K(\rho)} \sigma^*(z|\rho)g(z|\rho)dz} d\kappa \\ \mathcal{F}(\rho) &= \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} \int_{K(\rho)} \frac{\sigma^*(\kappa, \rho)g(\kappa|\rho)}{\int_{K(\rho)} \sigma^*(z|\rho)g(z|\rho)dz} d\kappa \\ \mathcal{F}(\rho) &= \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho}\end{aligned}$$

This implies that

$$\mathcal{F}(B) = \frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1) - \delta(0))k_\rho} = \mathcal{F}(W) \quad (9)$$

Therefore, differential fatality rates can only arise if $k_\rho \neq k_{-\rho}$, which, by Definition 1 means O is racially biased. \square

Proof of Proposition 3. Notice that Condition (8) characterizes which types of civilians are willing to engage in potentially suspicious behavior and therefore be candidates for law-enforcement activity. That condition can be re-written as

$$d[\pi^*(\tau)(\delta(1) - \delta(0)) + \delta(0)] \leq \frac{\pi^*(\tau)b(\tau)(1 - \sigma^*) - w(\tau)}{\sigma^*}$$

Note that $\frac{\partial \sigma^*}{\partial k_\rho} > 0$. Therefore, all else equal, as k_ρ decreases—as the officer becomes increasingly biased towards using lethal force against a civilian of race ρ , then this inequality is less likely to hold. \square

C Lower bound derivation for magnitude of racial bias

Given two groups, Black civilians and White civilians, where $\rho_i \in \{b, w\}$ indicates civilian i 's race, and the decision to shoot civilian i is defined as $S_i \in \{0, 1\}$. If $Y \in \{0, 1\}$ is the outcome, we can define the probability of police fatally shooting a civilian of race ρ as follows:

Definition 2. *The rate of of being fatally shot among a racial group, ρ , is given by $\mathcal{F}_\rho = E[Y|S(\rho) = 1, \rho]$.*

Per our model, let $\rho = b$ indicate a Black civilian. The fatality rate among Black civilians, \mathcal{F}_b , can be expressed as a combination of Black civilians who would have been shot had they been White ($\mathcal{F}_{s(b)=s(w)}$), and Black civilians who would *not* have been shot had they been White ($\mathcal{F}_{s(b)>s(w)}$). The magnitude of

racial bias in the decision to shoot a civilian is the proportion of Black civilians shot who would not have been shot had they been White, $p = \Pr[S(w) = 0, S(b) = 1 | \rho = b]$. This can be formally written as:

$$\mathcal{F}_b = p \cdot \mathcal{F}_{s(b) > s(w), b} + (1 - p) \cdot \mathcal{F}_{s(b) = s(w), b} \quad (10)$$

Our quantity of interest is p from Equation (10)—the *proportion of Black civilians who were shot (both fatally and non-fatally) who would not have been shot were they White*. In order to calculate a lower bound on p , we first assume monotonicity in the direction of racial bias. Specifically, we assume there are no White civilians that would not have been shot had they been Black. That is, racial bias does not lead officers to decline to shoot civilians simply because they are Black. The important assumption is therefore that there are not fatal shootings among White Civilians who would not have been shot had they been Black— $\mathcal{F}_{s(w) > s(b), w} = E[Y | S(w) = 1, S(b) = 0, \rho = w] = 0$.

Assumption 3 (Monotonicity). *The probability that a White civilian is fatally shot who would not have been fatally shot had she been Black is zero. Formally, $\Pr[S(w) = 1, S(b) = 0] = 0$. This implies $E[Y | \rho_i = w] = E[Y | S(w) = 1, S(b) = 1, \rho = w]$, which can be written as $\mathcal{F}_w = \mathcal{F}_{s(w) = s(b), w}$.*

Assumption 3 leaves three principle strata for definition—the fatality rate among Black civilians who would not have been shot were they White (i.e., people who were shot because they were Black), the fatality rate among Black civilians who would have been shot were they White (i.e., independent of their race), and the fatality rate among White civilians who would have been shot were they Black. As in Assumption 3, we define these quantities by expected outcomes for each situation.

Definition 3. *The mean outcome among those shot with $\rho = b$, that would not have been shot if they had been $\rho = w$, is given by $\mathcal{F}_{s(b) > s(w), b} = E[Y | S(w) = 0, S(b) = 1, \rho = b]$. The mean outcome among those shot with $\rho = b$ that would have been shot if they had been $\rho = w$, is given by $\mathcal{F}_{s(b) = s(w)} = E[Y | S(w) = 1, S(b) = 1, \rho = b]$. The mean outcome among those shot with $\rho = w$ that would have been shot if they had been $\rho = b$ $\mathcal{F}_{s(w) = s(b)} = E[Y | S(w) = 1, S(b) = 1, \rho = w]$.*

We now use the monotonicity assumption in conjunction with our observed data to calculate the proportion of Black civilians shot in our nine localities who would not have been shot had they been White civilians. We can rearrange Equation (10) to solve for p :

$$p = \frac{\mathcal{F}_b - \mathcal{F}_{s(b) = s(w), b}}{\mathcal{F}_{s(b) > s(w), b} - \mathcal{F}_{s(b) = s(w), b}}.$$

Notice, we observe both \mathcal{F}_b , the rate of being fatally shot among all Black civilians shot, and \mathcal{F}_w , the rate of being fatally shot among all White civilians shot. Also, given Assumption 3 and Definition 3, we can substitute \mathcal{F}_w for $\mathcal{F}_{s(b) = s(w), w}$ —the fatality rate among all White civilians shot can stand in for the fatality rate for Black civilians who would have been shot were they White. Note that the logic of our formal analysis

and our empirical findings both indicate $\mathcal{F}_w > \mathcal{F}_b$. Therefore, we rewrite this expression of p :

$$p = \frac{\mathcal{F}_w - \mathcal{F}_b}{\mathcal{F}_w - \mathcal{F}_{s(b) > s(w), b}}. \quad (11)$$

Our p is expressed as the difference between the observed rate of being fatally shot among White and Black civilians, divided by the difference between the observed rate of being fatally shot among White civilians and the rate of being fatally shot among Black civilians who would not have been shot were they White, $\mathcal{F}_{s(b) > s(w)}$. That quantity is not observed because we do not know precisely who would not have been fatally shot had they been White. However, if there is no racial bias, we know that $\mathcal{F}_{s(b) > s(w)}$ would be 0, as there would be no civilians shot because they were Black. Therefore, substituting 0 for $\mathcal{F}_{s(b) > s(w)}$ yields a lower bound on the true value of p :

$$p \geq \frac{\mathcal{F}_w - \mathcal{F}_b}{\mathcal{F}_w - 0} = \frac{\mathcal{F}_w - \mathcal{F}_b}{\mathcal{F}_w}. \quad (12)$$

To estimate the lower bound on the proportion of Black civilians that would not have been shot had they been White, we begin by estimating a linear probability model. We estimate it with a subset of officer-involved shooting data containing only Black and White civilians, including our main covariate of interest, namely race (White equal to 1, Black equal to 0), along with binary indicator variables for locality. Using this model we estimate the regression coefficient on White to be 0.17 (see Appendix Table 10 for the full regression specification results), the associated fatality difference between White civilians and Black civilians controlling for city fixed effects. Equation 13 defines the regression-based lower bound estimate.

$$p \geq \frac{\beta_w}{F_b + \beta_w} \quad (13)$$

The coefficient estimate on White is the numerator and the coefficient plus the observed mean fatality rate among Black civilians is the denominator. Thus, we estimate 33.3% (see Table 5) is the lower bound on the proportion of Black civilians that police would not have shot had they been White.

Lower Bound Estimate	Confidence Interval around Lower Bound
0.333	[0.219,1]

Table 5: *Lower bound estimate using regression.* Lower confidence interval around the lower bound is estimated by bootstrapping. The upper interval is necessarily 1 because it could be that 100% of Black civilians would not have been shot had they been White.

Substantively, our estimate of 33.3% is considerable and given it is a lower bound, may be much higher.

Our estimate implies that police would not have shot 166 Black civilians had they been White, from the 497 Black civilians in our nine localities over the years we study. Extrapolating this estimate to the larger population of the United States, however, is beyond the limits of our data. Moreover, significant intra-locality variation suggests police behavior, measured by officer-involved shootings, is not uniform across the country. Additionally, comparing Hispanic civilians and Asian civilians to White civilians yielded no statistically significant differences. That is consistent with what we would expect—police officers differentially exercise discretion against Black civilians as compared to all other groups. Given the extant debate about whether the use of force by police is tainted with racial bias, these findings suggest there is a substantively significant problem. Quantifying the magnitude of its effect, though, requires richer administrative data beyond what police departments, generally, in the U.S. currently provide. Specifically, the important matter of how much police violence is attributable to racial bias requires knowing how often police fire their weapons, as well as how often they draw their weapons (John L. Worrall et al., 2018; Wheeler et al., 2017), which is not universally known across local police departments.

D Assessing the Mechanism

We argue the mechanism at work in the empirical evidence we have shown is that racially biased officers have a lower threshold for using force against racial minorities than against White civilians. The evidence we have shown is consistent with the consequences of such bias. We now step back to assess broader evidence, outside the context of officer-involved shootings, to corroborate our claim about the underlying mechanism. In particular, we consider whether we do in fact observe lower thresholds for using force by when officers encounter Black civilians, as compared to White civilians.

To do so, we marshal several related datasets on officer-civilian interactions and show a consistent pattern across a variety of jurisdictions. We note at the outset, these data comprise incidents of officer-civilian interactions that were recorded and so suffer from problems of selection bias and unobservable counterfactuals. However, our goal here is to demonstrate there are patterns beyond those we have documented that are consistent the claim that officers have a lower threshold for using force against Black civilians than against White civilians. To the extent we find evidence consistent with that claim, we can be more confident that the patterns in civilian fatalities in officer-involved shootings are caused by the theoretical model we have proposed, as opposed to some other process.

Table 6 summarizes the data we have assembled. From New York City, we have the widely studied Stop, Question, and Frisk data, which comprise 11 years of data on incidents in which officers stop civilians and contain detailed information about actions taken by the officer during the encounter. One very widely

Jurisdiction	Years covered	Brief description
New York City	2006-2015	Stops with indicators for different kinds of force
Washington, DC	4 weeks in 2019	Data on all stops, outcome is whether a pat-down was conducted

Table 6: *Summary of civilian contact data used to assess racial disparities in the use of force.*

Dependent variable:	Washington, DC	NYC
	Pat-down	Force Used
Black Civilian	0.09*** (0.01)	0.04*** (< 0.01)
Hispanic Civilian	0.01 (0.01)	
Black-Hispanic Civilian		0.04*** (< 0.01)
White-Hispanic Civilian		0.02*** (< 0.01)
Asian Civilian	0.00 (0.02)	-0.01*** (< 0.01)
Native American		0.00 (< 0.01)
Multiple Race Civilian	-0.01 (0.05)	
Other Race Civilian	-0.02 (0.14)	0.01*** (< 0.01)
Unknown Civilian Race	-0.01 (0.02)	-0.06*** (0.01)
N		5029789
Fixed effects	Date, district	Month-year, precinct

Table 7: *Racial disparities in the use of force in selected datasets.* Entries are linear regression coefficients, standard errors in parentheses. *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

studied source of variation in these data are indicators for different kinds of force that an officer may have used during a stop. From Washington, DC, we have a recently-released dataset that comprises just four weeks of stops during 2019 but include an indicator for whether an officer conducted a pat-down of the civilian involved in the stop.

Table 7 reports the results of a series of fixed effects linear regression models in which the dependent variables are indicators of force or the conducting of a pat-down. The explanatory variables are fixed effects for civilian race as well as other fixed effects, depending on what is available and feasible from each city. For example, we have more than 5,000,000 observations from New York, and so we include fixed effects for every month-year pair during our window, as well as the precinct in which the stop took place. For Washington, DC, we only have a few weeks' data, and so we include date-specific fixed effects, along with the district in which the stop took place. In each of the models we specify, we use White civilians as the baseline category, so the table entries can be interpreted as differences between the groups in the table and White civilians.

Across all of our specifications, Black civilians are more likely to be subjected to force than are White

civilians. These data are consistent with the theoretical mechanism underlying our model—that officers might have a lower threshold for using force against a Black civilian than a White civilian. We caution, though, these data are less closely connected to our model and so should be interpreted with caution. However, we believe they do provide at least preliminary additional evidence of the theoretical mechanism we contemplate.

E Additional Model Specifications

Table 8 repeats the original specification from the main paper with city fixed effects as Model 1. The additional specifications include year fixed effects. Notably, the magnitude of the relationship between being a Black civilian and the probability of dying *increases* once we include jurisdiction fixed-effects, and maintains when we include year fixed-effects.

	Model 1	Model 2	Model 3	Model 4
Black	−0.697*** (0.170)	−0.839*** (0.162)	−0.767*** (0.173)	−0.755*** (0.191)
Hispanic	0.074 (0.174)	0.259 (0.157)	0.055 (0.175)	0.071 (0.184)
Asian/AI/AN/PI	0.806* (0.389)	0.985** (0.380)	0.830* (0.395)	0.874* (0.438)
Houston	0.008 (0.412)		0.063 (0.415)	1.313 (1.132)
King County	0.270 (0.548)		0.296 (0.555)	0.298 (2.004)
Los Angeles	1.274** (0.396)		1.324*** (0.400)	2.889** (1.081)
Orlando	0.589 (0.453)		0.586 (0.455)	2.166 (1.221)
San Antonio	1.786** (0.660)		1.936** (0.682)	2.620 (2.082)
San Jose	0.197 (0.506)		0.231 (0.511)	1.156 (1.562)
Seattle	1.252** (0.449)		1.186** (0.450)	2.322 (1.216)
Tucson	1.621*** (0.456)		1.544*** (0.465)	17.000 (1455.398)
2011		0.446* (0.227)	0.190 (0.239)	−12.503 (1029.122)
2012		0.198 (0.233)	0.140 (0.244)	3.250* (1.629)
2013		0.318 (0.227)	0.171 (0.236)	2.131 (1.264)
2014		0.566* (0.240)	0.440 (0.250)	0.983 (1.499)
2015		0.171 (0.247)	0.127 (0.257)	3.063* (1.457)
2016		0.215 (0.266)	0.101 (0.282)	1.323 (1.076)
2017		−0.092 (0.235)	−0.212 (0.247)	1.239 (1.420)
Intercept	−0.800* (0.398)	−0.134 (0.208)	−0.920* (0.418)	−2.308* (1.047)
City*Year				✓
N	1292	1292	1292	1292
AIC	1644.834	1720.182	1651.172	1654.931

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 8: Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show logistic coefficients with standard errors. Omitted category is White civilians, Charlotte and the year 2010.

Table 9 includes linear probability models. In each of our specifications, among those civilians shot by an officer, Black civilians are less likely to die than are White civilians. This difference is statistically significant in each specification. Notably, as with the logistic model discussed in the paper, the magnitude of the relationship between being a Black civilian and the probability of dying *increases* once we include jurisdiction fixed-effects, and maintains when we include year fixed-effects. What is more, the relationship between being a Hispanic civilian and a reduced probability of dying does not emerge even after we include jurisdiction-level and year fixed effects. This functions as a placebo test and implies that any problematic unmeasured covariates would have to have different relationships for Black and Hispanic civilians (e.g., concerns about characteristics that affect the probability of death—such as police behavior, training, and medical attention would be largely ruled out by this analysis).

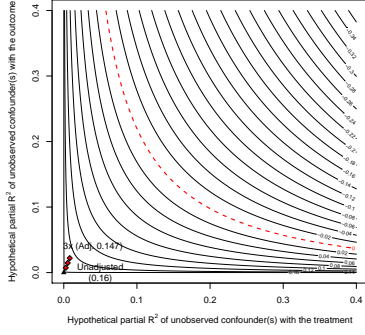
F Sensitivity Analysis

Using the specification from Model 5 in Table 9 we conducted sensitivity analysis based on Cinelli and Hazlett, 2020. This sensitivity analysis works by considering how strong an unmeasured confounding variable would have to be in order to wipe out the effects we are finding for Black civilians. One method Cinelli and Hazlett, 2020 provides to measure such strength is to benchmark any potential unmeasured confounder against measured covariates in the model. Each contour plot uses a different benchmark variable to show that any such unmeasured confounding variable would have to be more than three times as strong as any of variables we currently have in the model (jurisdiction fixed effects, time fixed effects, and distance to trauma center). The red dashed contour line in each plot shows how strong an unmeasured confounder would have to be to wipe out the effect for Black civilians. The adjusted estimates show how the covariate for Black would change with an unmeasured confounder 1, 2 and 3 times as large as the benchmarked variable. All of these are well below the red dashed contour line showing no effect. We cannot think of any such missing variable. We especially cannot think of any such missing variable that would affect fatality rates for Black civilians and not Hispanic civilians.

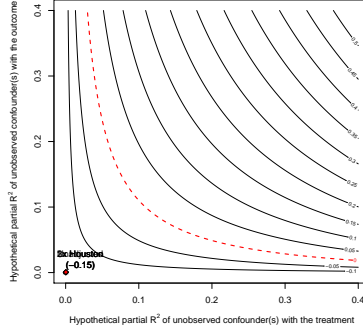
	Model 1	Model 2	Model 3	Model 4	Model 5
Black	−0.156*** (0.038)	−0.200*** (0.038)	−0.170*** (0.038)	−0.157*** (0.040)	−0.150*** (0.038)
Hispanic	0.019 (0.039)	0.063 (0.038)	0.015 (0.039)	0.021 (0.040)	0.026 (0.039)
Asian/AI/AN/PI	0.174* (0.080)	0.225** (0.083)	0.178* (0.081)	0.169* (0.086)	0.192* (0.080)
Distance					0.009*** (0.003)
Houston	−0.003 (0.080)		0.013 (0.081)	0.180 (0.157)	−0.048 (0.086)
King County	0.048 (0.113)		0.055 (0.115)	−0.057 (0.390)	−0.032 (0.119)
Los Angeles	0.281*** (0.078)		0.293*** (0.079)	0.519*** (0.147)	0.257** (0.083)
Orlando	0.115 (0.091)		0.117 (0.091)	0.343 (0.191)	0.100 (0.096)
San Antonio	0.395** (0.135)		0.430** (0.140)	0.427 (0.380)	0.409** (0.143)
San Jose	0.026 (0.105)		0.036 (0.105)	0.126 (0.269)	0.021 (0.108)
Seattle	0.274** (0.092)		0.262** (0.092)	0.415* (0.191)	0.278** (0.096)
Tucson	0.361*** (0.092)		0.345*** (0.094)	0.723 (0.490)	0.367*** (0.096)
2011		0.104* (0.053)	0.043 (0.053)	0.049 (0.356)	
2012		0.045 (0.054)	0.034 (0.053)	0.604* (0.300)	
2013		0.073 (0.053)	0.039 (0.051)	0.333 (0.200)	
2014		0.131* (0.056)	0.097 (0.055)	0.160 (0.207)	
2015		0.039 (0.057)	0.030 (0.056)	0.549* (0.270)	
2016		0.049 (0.062)	0.021 (0.062)	0.275 (0.234)	
2017		−0.024 (0.055)	−0.042 (0.054)	0.281 (0.299)	
Intercept	0.329*** (0.079)	0.471*** (0.049)	0.300*** (0.084)	0.108 (0.134)	0.286*** (0.084)
N	1292	1292	1292	1292	1266
RMSE	0.470	0.484	0.470	0.464	0.469

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

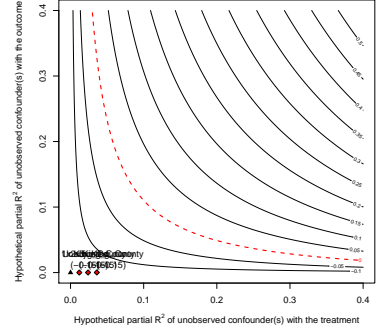
Table 9: Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show linear model coefficients with standard errors. Omitted category is White civilians, Charlotte and the year 2010.



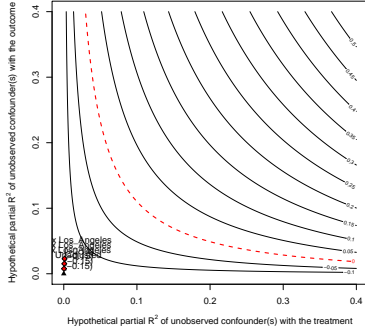
(a) Benchmark: Distance



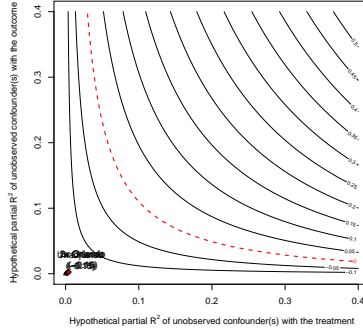
(b) Benchmark: Houston



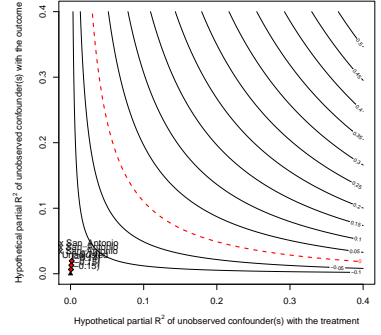
(c) Benchmark: King County



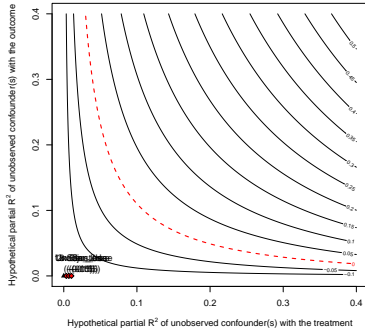
(d) Benchmark: Los Angeles



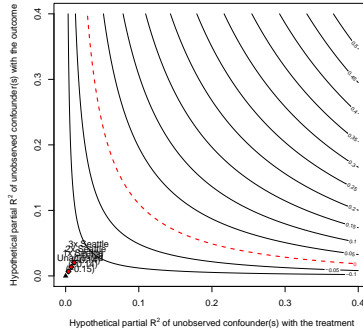
(e) Benchmark: Orlando



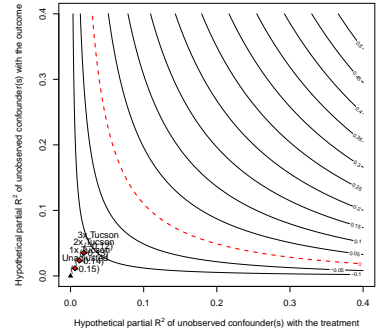
(f) Benchmark: San Antonio



(g) Benchmark: San Jose

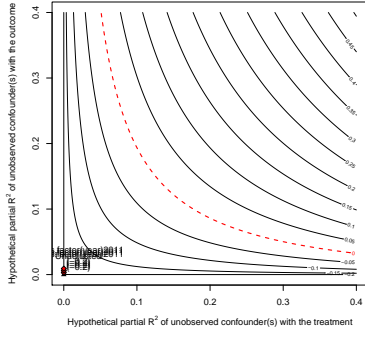


(h) Benchmark: Seattle

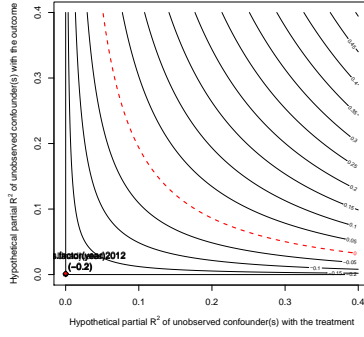


(i) Benchmark: Tucson

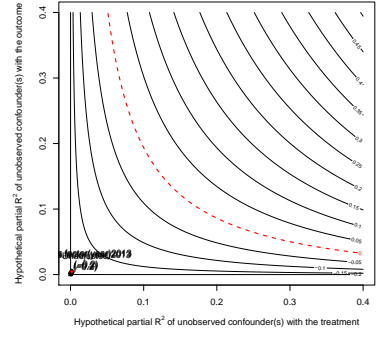
Figure 4: Sensitivity analysis using distance to trauma center and jurisdiction fixed effects as benchmarks



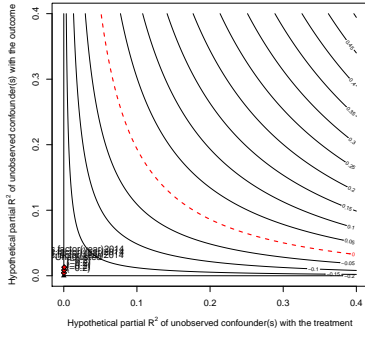
(a) Benchmark: Year 2011



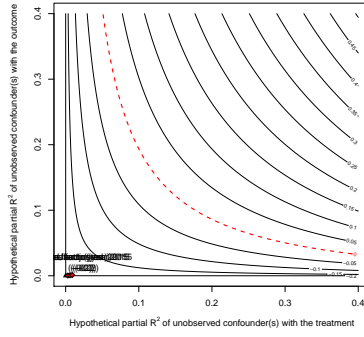
(b) Benchmark: Year 2012



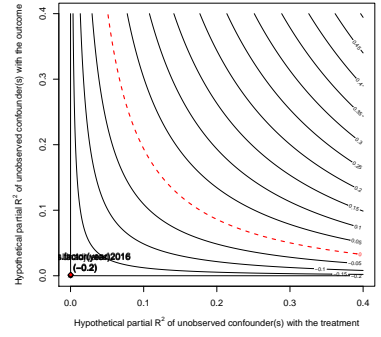
(c) Benchmark: Year 2013



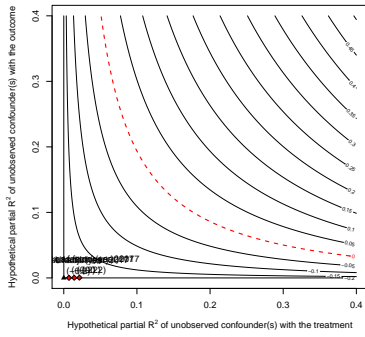
(d) Benchmark: Year 2014



(e) Benchmark: Year 2015



(f) Benchmark: Year 2016



(g) Benchmark: Year 2017

Figure 5: Sensitivity analysis using year fixed effects as benchmarks

F.1 Sensitivity Analysis for Outcome Test

To estimate the lower bound on the proportion of Black civilians that would not have been shot had they been White, we begin by estimating a linear probability model. Table 10 shows the regression results for three different subsets of the data — Black and White civilians, Hispanic and White civilians and Asian and White civilians. Our main results in the paper focus on the subset of officer-involved shooting data containing only Black and White civilians, including our main covariate of interest, namely race (White equal to 1, Black equal to 0), along with binary indicator variables for locality. Using this model we estimate the regression coefficient on White to be 0.17 the associated fatality difference between White civilians and Black civilians controlling for city fixed effects. We see that neither Hispanic nor Asian civilians have a statistically significant associated fatality difference with White civilians controlling for city fixed effects.

DV: Data:	Fatal (Black/White)	Fatal (Hispanic/White)	Fatal (Asian/White)
White	0.17*** (0.04)	−0.04 (0.04)	−0.14 (0.08)
Charlotte	0.16* (0.07)	0.22 (0.12)	0.37* (0.15)
Houston	0.20*** (0.04)	0.31*** (0.05)	0.37*** (0.10)
King_County	0.24* (0.10)	0.54*** (0.12)	0.63*** (0.12)
Los_Angeles	0.44*** (0.03)	0.63*** (0.02)	0.73*** (0.08)
Orlando	0.30*** (0.05)	0.37** (0.12)	0.47*** (0.14)
San_Antonio	0.92*** (0.24)	0.69*** (0.12)	1.14*** (0.34)
San_Jose	0.25* (0.12)	0.40*** (0.08)	0.54*** (0.15)
Seattle	0.43*** (0.06)	0.77*** (0.09)	0.89*** (0.09)
Tucson	0.40*** (0.07)	0.81*** (0.06)	0.86*** (0.11)
R ²	0.45	0.61	0.62
Adj. R ²	0.44	0.60	0.61
Num. obs.	746	762	290
RMSE	0.47	0.47	0.46

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Linear probability model regression results. Each model is run on a subset of the data to make either a Black/White, Hispanic/White or Asian/White comparison.

References

- Alpert, Geoffrey P. and Roger G. Dunham (2004). *Understanding Police Use of Force: Officers, Suspects, and Reciprocity*. Cambridge University Press.
- Belvedere, Kimberly, John L Worrall, and Stephen G Tibbetts (2005). “Explaining suspect resistance in police-citizen encounters”. In: *Criminal Justice Review* 30.1, pp. 30–44.
- Bierie, David M (2017). “Assault of police”. In: *Crime & Delinquency* 63.8, pp. 899–925.
- Bierie, David M, Paul J Detar, and Sarah W Craun (2016). “Firearm violence directed at police”. In: *Crime & Delinquency* 62.4, pp. 501–524.
- Cinelli, Carlos and Chad Hazlett (2020). “Making sense of sensitivity: Extending omitted variable bias”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67.
- Kahn, Kimberly Barsamian et al. (2017). “How suspect race affects police use of force in an interaction over time.” In: *Law and human behavior* 41.2, p. 117.
- Knox, Dean, Will Lowe, and Jonathan Mummolo (forthcoming). “Administrative Records Mask Racially Biased Policing”. In: *American Political Science Review*.
- Wheeler, Andrew P. et al. (2017). “What Factors Influence an Officer’s Decision to Shoot? The Promise and Limitations of Using Public Data”. In: *Justice Research and Policy* 18.1, pp. 48–76.
- Worrall, John L. et al. (2018). “Exploring Bias in Police Shooting Decisions With Real Shoot/Don’t Shoot Cases”. In: *Crime & Delinquency*, pp. 1171–1192.