Are Police Racially Biased in the Decision to Shoot?*

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

January 18, 2021

Abstract

We present a theoretical model predicting racially biased policing produces 1) more use of potentially lethal force by firearms against Black civilians than against White civilians and 2) lower fatality rates for Black civilians than White civilians. We empirically evaluate this second prediction with original officer-involved shooting data from nine local police jurisdictions from 2005 to 2017, finding that Black fatality rates are significantly lower than White fatality rates, conditional upon civilians being shot by the police. Using outcome test methodology, we estimate that at least 30% of Black civilians shot by the police would not have been shot had they been White. We also show that an omitted covariate three times stronger than our strongest included covariate would only reduce this estimate to 18%. Additionally, such an omitted covariate would have to affect Black fatality rates and not Hispanic fatality rates in order to be consistent with the data.

*We thank Chuck Cameron, Anna Harvey, Hakeem Jefferson, John Kastellec, Beatriz Magaloni-Kerpel, Pablo Montagnes, John Patty, Zac Peskowitz, Dan Thompson, Georg Vanberg, Kirsten Widner, and seminar participants at Emory University, Princeton University, the University of California-San Diego, and the Institute for Advanced Study in Toulouse 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

Police are agents of the state, exercising a high degree of autonomy and discretion when implementing policy (M. K. Brown, 1981; Wilson, 1978). Unlike other domestic agents of the state, however, "the police are...a mechanism for the distribution of situationally justified force in society" (Bittner, 1970, p. 39). Consequently, the character of their interactions with the public differ greatly from those of other "street-level bureaucrats": Police-civilian encounters are more unpredictable, with greater potential for violence and death, for civilians and police. Accordingly, policing is "profoundly involved with the most significant questions facing any political order, those pertaining to justice, order, and equity" (M. K. Brown, 1981, pp. 6-7). It is especially true when police use their discretion to shoot civilians.

While police use force against civilians more 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, racial disparities in police use of force in the U.S. seem common and particularly wide between Blacks and Whites, and a marker of racial disparities in policing, generally, including deployment, surveillance, involuntary contact by stop-and-frisk, arrest, and jailing (Bittner, 1970; Soss and Weaver, 2017; R. A. Brown, 2019). Given the fraught history and contemporary realities of race in the U.S., racial disparity in police shootings raise concerns about racial bias influencing officers' discretion to shoot during police-civilian encounters. Whether racial bias causes racial disparities in policing, and how much, however, remains an academic and civic puzzle.

It is empirically difficult to discern how many police shootings of Black Americans result from their disproportionate contact with police versus disproportionate use of force by police against them versus racial bias by patrol officers and their departments (e.g., Knox, Lowe, and Mummolo, 2020; Knowles, Persico, and Todd, 2001; Fryer Jr., 2016). Further, neither police departments nor agencies overseeing them 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 differences in fatal shootings of civilians by police (Johnson et al., 2019). Therefore, even when relatively good data are available for social scientists to observe and describe racial patterns in policing, scholarly consensus on whether and how much 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 choose their behavior. Behavior may include actions that escalate their encounters towards harm, including police violence against civilians (and civilian violence against police). Our model predicts that racially biased police officers will be more likely to use force against Black civilians than against White civilians. Moreover, police shootings of Black civilians should result in more non-fatalities than fatalities, as we will explain and empirically show.

We test the implication of our model with OIS data from nine local police jurisdictions in the U.S.. Our data, obtained through public records requests, covering 2005 through 2017, include all instances of civilians shot by local police—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, 2020 and Cohen, 2020. Borrowing these techniques, 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 their White fatality rates, 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. Using the techniques from Cohen, 2020, we estimate that at least 30% of Black civilians shot would not have been shot had they been White. We also show that an omitted covariate three times stronger than our strongest included covariate would only reduce this estimate to 18%. Additionally, such an omitted covariate would have to affect Black fatality rates and not Hispanic fatality rates in order to be consistent with the data. [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, 2020). That alone is important in light of the continuing need to understand discretion by the police as "street-level bureaucrats" (Lipsky, 1980) and how much race affects policing, including use and severity of force. 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 (M. K. Brown, 1981).

2 Police Discretion in Use of Force

Encounters with the police are among the most common encounters civilians have with government agents (Jacob, 1972; M. K. Brown, 1981; Soss and Weaver, 2017). A key contrast with other civilian encounters with government agents is that civilian-police contact, whether

[1] We lack data on all instances of police drawing their weapons, but including moments where police drew guns without firing would likely increase the estimate of the lower bound (Worrall et al., 2018).

initiated by police or initiated by civilians, has the potential for violence. How officers exercise their discretion to use force and violence during police-civilian encounters and why it may cause racial disparities are important considerations (e.g., Terrill, 2011).

2.1 Racial Disparities in Use of Force

Generally, social scientists expect 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 do is well-studied 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 police-civilian encounters at incident-level (e.g., New York City's Stop, Question, and Frisk program) has enabled rigorous social science to deepen evidence of racial disparities in police use of force (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).

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). In other words, racial bias in policing may not necessarily increase the likelihood of use of force against Black civilians. Some evidence, drawn typically from observational studies, 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 smaller-scale or no racial disparities in police use force (e.g., shootings). 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 that 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.

Different conceptions of racial bias can exist. The aforementioned quote focuses on the bias of the patrol officer that shoots. But, one might focus on the police department (and supervisors) of the officer. As Bittner, p. 10 posited, "The ecological deployment of police work at the level of departmentally determined concentrations of deployment, as well as in terms of the orientations of individual police officers, reflects a whole range of public prejudices." We focus on bias by the patrol officer, recognizing the importance of administrative control and bureaucratic bias (M. K. Brown, 1981).

Additionally, 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 police 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 require accounting for confounds affecting contact and use of force (Knox, Lowe, and Mummolo, 2020). Race, for example, may be correlated with other characteristics (e.g., income, education, geography,

[2] By "race" of civilian, we mean the officer's perception of their race.

employment, social networks) that might cause disparate rates of contact with police, thereby influencing civilian exposure to police use of force. Therefore, racially disparate patterns in the use of force and its severity may spuriously relate to characteristics of civilian-officer encounters that explain use of force (e.g., Jetelina et al., 2017; Worrall et al., 2018; Knowles, Persico, and Todd, 2001; Cesario, Johnson, and Terrill, 2019). To best study the effect of race on the propensities of civilians to experience police use of force requires conditioning on a range of civilian characteristics that may confound the relationship. Furthermore, there is the matter of selection into contact with police and how it challenges inference-drawing about racial bias during citizen-police interactions (Johnson et al., 2019; Knox and Mummolo, 2019; Knox, Lowe, and Mummolo, 2020).

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 police using greater force against them: They are more likely to come into contact with police because police officers racially profile them[3] or they are more proximate to high-crime, highly-policed environments. The second mechanism is that police officers differentially perceive the stakes for using force against civilians depending on the race of the civilians. Officers might, for example, anticipate differential downstream consequences from using force against Black civilians than from using force against White civilians. 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.

3 A Racial Bias Model of Police Shootings

Our racial bias model of police shootings stems from the model Knowles, Persico, and Todd (2001) employ to examine police stops of drivers. Also, our model seeks to capture "the transactional, or step-by-step unfolding, of police–public encounters" and the "micro process

[3]Racial profiling as a mechanism of racial disparities in use of force, however, is potentially circular.

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. It allows for disparate rates of encounters with the police across racial groups. Modeling this stage allows us to make empirical predictions about behavior *implied* by racial bias that should manifest even in the presence of strategic selection into encounters 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 violence. In the second stage, we model a conflict subgame that seeks to capture the kinds of 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 civilian have a gun or a wallet) suggest this process is accurately captured by simultaneous structure.

In our model, conflict takes the form of aggression 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.).[4] *We assume that the likelihood of death following police shooting a civilian is, in part, affected by the civilian's behavior.* To be clear, we use "aggression" in a specific way—civilian behavior perceived by police as defiant or belligerent that enhances the likelihood of death when the officer uses force. We assume that during police-civilian encounters that become violent, the officer is advantaged by firepower and manpower, training and expertise. While testing this assumption is beyond the scope of our analysis, we believe it is at least plausible. Indeed, empirical studies provide reason to believe that threatening and/or resisting an officer or otherwise escalating a situation through perceived aggression increases the likelihood that,

[4]We conceive of threatening behavior by the civilian broadly—it includes posing a "threat" to the officer, escalation in aggression, or other actions that heighten the stakes of conflict between the civilian and the officer. The degree of civilian non-compliance with police directives, in particular, accounts for much of the variance in use of force by police against civilians (e.g., J. H. Garner, C. D. Maxwell, and Heraux, 2002).

conditional upon being in a shooting, a civilian dies (e.g., Mohandie, Meloy, and Collins, 2009). What is more, the foundations of de-escalation training for police rest, in part, on the presumption that safety (for both the civilian and the officer) increases with de-escalation.

We assume officers and civilians have choice over their behavior and prefer physical safety. In particular, we model the possibility of racial bias by allowing officer perceptions of the cost of fatally shooting a civilian to vary by race of civilian. Our formal representation captures emotional reactions, anxiety and threat perception associated with racial bias and the use of force (e.g. Kleider, Parrott, and King, 2010; Nieuwenhuys, Savelsbergh, and Oudejans, 2012; Welch, 2007; Correll, Park, Judd, and Wittenbrink, 2002), along with a more dispassionate cost-benefit analysis by the officer about the *anticipated* consequences of killing a civilian.

3.1 Primitives

Players, sequence of play, and strategies. The model is played between a civilian, C, and an 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 observable civilian characteristics, denoted $\kappa \in \mathbb{R}$. The latter include dress, demeanor, location, time, 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 any racial group. 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 the interaction is drawn.

The sequence of play is summarized in Figure 1. 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" 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, by an officer (or another civilian). If the "suspect" civilian chooses $s = 0$, the game ends. However, if the "suspect" civilian chooses $s = 1$, then the officer must use their discretion to decide whether to engage the civilian, for purposes of order maintenance or law enforcement (e.g., stop-question-frisk). Let $l \in \{0, 1\}$ denote this choice, with $l = 1$ denoting engaging the civilian. If the officer chooses $l = 0$, the game ends; if he chooses $l = 1$, the game proceeds to the next stage, with simultaneous interactions by civilian and officer. Specifically, both players must decide whether to engage escalate aggression. The civilian must choose 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 lethal force. If the officer chooses 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 escalates 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

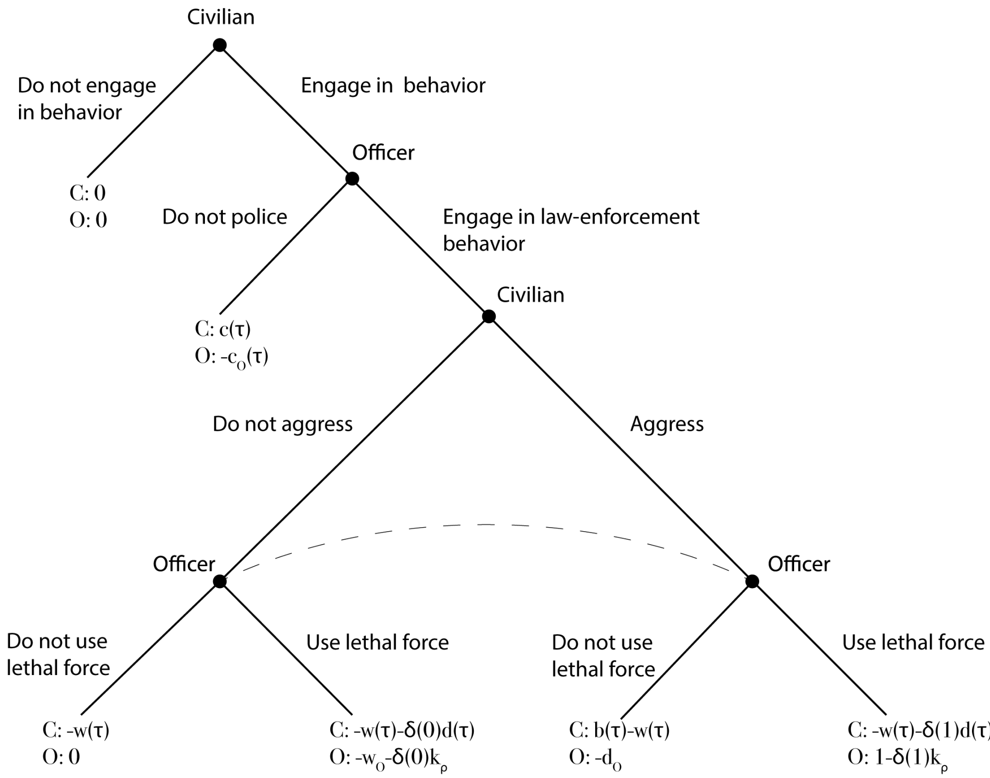


Figure 1: *Sequence of play in the model.*

lethal force or not, as well as whether the civilian chooses a behavior that escalates conflict. 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 policing 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 policing, 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 the loss of life, which can depend on civilian type—i.e., some civilians may value living more than others such as the suicidal. To avoid considering unreasonable situations, we assume that the cost of dying is worse than the cost of being subjected to policing 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 chooses less-than-lethal force, 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 given by:

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

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 civilian type, allowing an officer's disutility from permitting potentially criminal activity to occur 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. By contrast, the officer pays a cost, $-d_O$, where $d_O > 0$ whenever a civilian is aggressive 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 another adverse consequence. Finally, we assume the officer receives positive utility 1 from using force to stop a civilian who is engaged in criminal activity and acting 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:

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

3.2 Analysis

We characterize a mixed-strategy subgame perfect Nash equilibrium. There can exist a pure strategy equilibrium if officers are never willing to use lethal force, which we rule implausible by assumption. 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. There is a probability that satisfies this requirement:

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

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 aggressive behavior. 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:

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

Thus, in any equilibrium that reaches the conflict subgame, there exists a mixed-strategy subgame perfect Nash equilibrium where civilians probabilistically aggress and officers probabilistically use lethal force.[5]

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

3.3 Empirical Implications

How does racial bias by police officers 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 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

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

who are shot is given by

$$\mathcal{F}(\rho) = \int\limits_{K(\rho)} \left(\delta(1)\cdot\pi^*(\tau) + \delta(0)\cdot(1-\pi^*(\tau))\right) \frac{\sigma^*(\tau)\, g(\kappa|\rho)}{\int\limits_{K(\rho)} \sigma^*(z|\rho) g(z|\rho) dz} d\kappa \tag{3}$$

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

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

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 is that if police are not racially biased then the probability a civilian is killed in an OIS, conditional on being involved in a shooting, should be independent of her race, even accounting for 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 encounter. 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.

A corollary implication of our model is that White civilians should be more likely than Black civilians to engage in escalation and aggression towards officers. That implication is important, because it helps clarify the underlying theoretical mechanism we posit. Black civilians are induced to be more cautious during an interaction with police than are White civilians.

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.*

It is beyond the limits of this paper to fully investigate that implication, but its verisimilitude is important for establishing the mechanism that drives the analysis we present. To that end, we note that beyond anecdotal support for the mechanism, there is some evidence from extant literature to support the implication. Kavanagh (1997) studies more than 1,000 encounters between civilians and officers in New York City'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 the underlying mechanism we posit. However, for the remainder of the paper, we evaluate the primary implication of the mechanism articulated above.

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 local police jurisdictions. 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 the effect of civilian race on the decision to use force; it is designed to demonstrate evidence implied by any such bias. In the fourth step, therefore, 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 in our sample of localities.

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 lendees repay mortgages on time at higher rates than Whites lendees, 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. Importantly, in many traditional settings, hit-rate tests are used to evaluate the presence of a latent trait in order to uncover evidence of bias. In our setting, as in Knowles, Persico, and Todd (2001), the latent trait is actually a choice by another player. In Knowles et al., drivers strategically choose whether to carry contraband; in our model, civilians strategically decide whether to escalate a confrontation and behave aggressively towards the officer. Anticipating bias by officers, Black civilians will be less likely in equilibrium to escalate a confrontation than will be White civilians. That feature is a result, and not an assumption. The motivating assumption, as we noted above, is that the risk of death is higher during a confrontation involving aggression than during one not involving aggression by the civilian.

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. They typically rely on media reports and crowd-sourced data, 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 OIS data by filing public records requests with individual police departments.

We sent public records requests to police departments and sheriffs' offices in the 50 largest local jurisdictions in the U.S., measured by population. We requested records of every single instance of an officer discharging their weapon between 2005 and 2017. Most policing agencies did not provide racial information about civilians involved in 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 response to our public records requests. The unit of analysis for each incident is the civilian/officer pair.

We constructed all civilian/officer pairs, yielding 1,292 total pairs, representing 774 unique incidents. Overall, 48% of our OIS incidents represent fatal shootings, varying considerably by department. San Antonio had the highest rate of fatal OIS 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. 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 jurisdictions. Because there is considerable variation in the size of the jurisdictions, there is considerable variation in the total number of

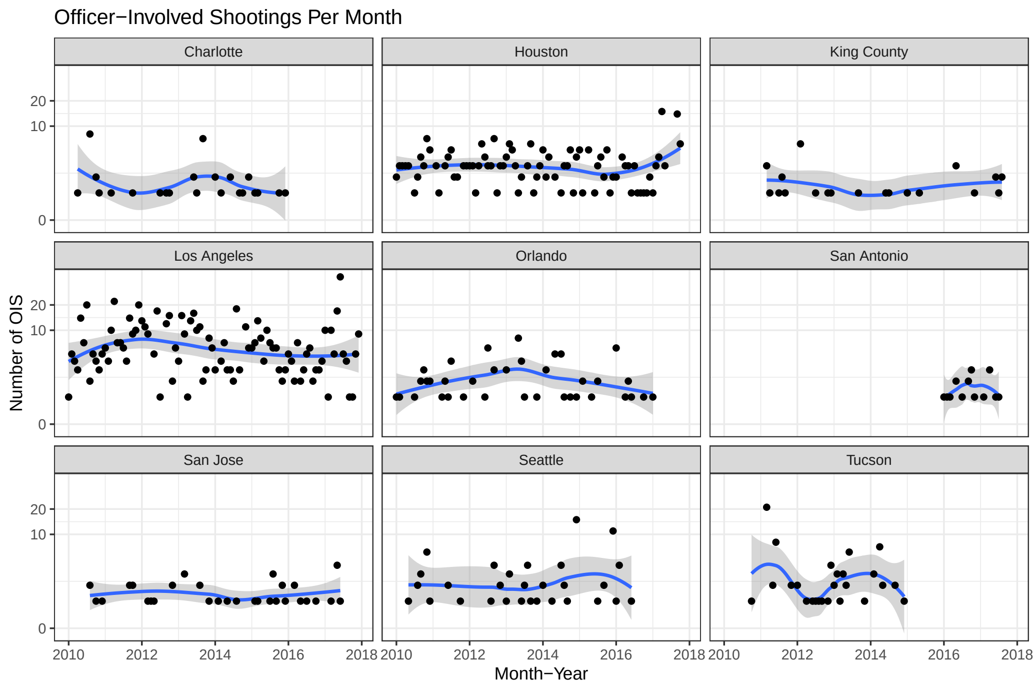
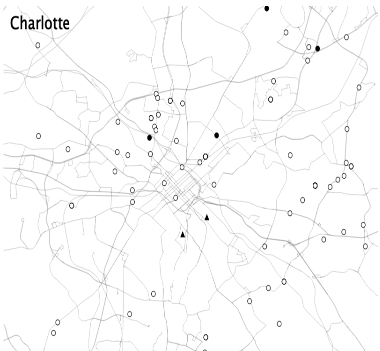


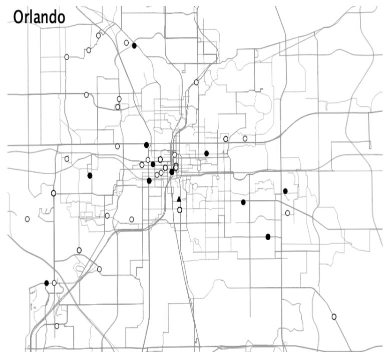
Figure 2: *Number of officer-involved shootings per month in nine cities, 2010-2017.* The figure plots the (logged) number of officer-involved shootings each month in each city.

OIS. The most come from Los Angeles, the second-largest police jurisdiction 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, there is little within-city variation in the frequency of OIS.

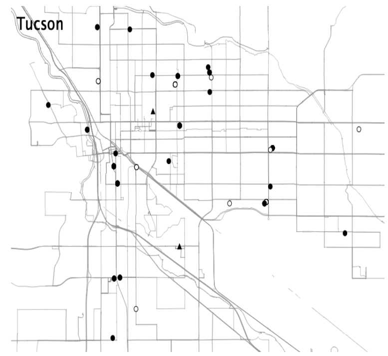
Furthermore, the spatial distribution and concentration of OIS within jurisdictions show intuitive but instructive patterns. Figure 3 shows the distribution of fatal and non-fatal shootings in our nine jurisdictions. Los Angeles and Houston, by far the largest localities in our dataset, experience the most OIS, whereas cities like Charlotte and Tucson experience relatively few. Additionally, it appears there is a higher fatality rate among OIS in localities like Los Angeles and Houston, which is less of an issue in jurisdictions 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 localities, and



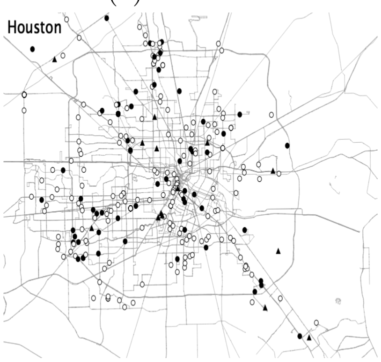
(a) Charlotte



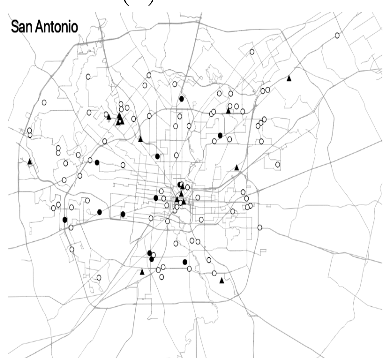
(b) Orlando



(c) Tucson

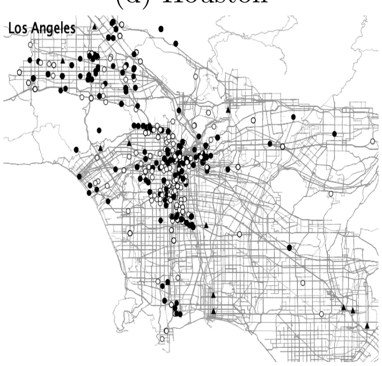


(d) Houston

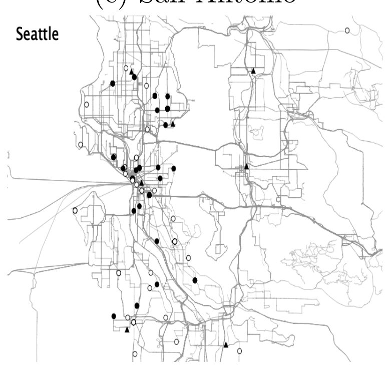


(e) San Antonio

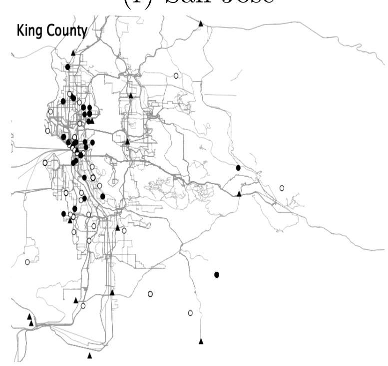
(f) San Jose



(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.

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 1 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 1: *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 1 is spurious. To speak to this we estimate a 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 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).[6] We also include fixed effects for

[6]Some observations lacked adequate location information to calculate the distance to the nearest trauma

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 1,292 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 2. 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

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^{***}	-0.70^{***}	-0.74^{***}	-0.67^{***}
	(0.16)	(0.17)	(0.16)	(0.17)
Hispanic	0.26	0.07	0.28	0.10
	(0.16)	(0.17)	(0.16)	(0.18)
Asian/AI/AN/PI	0.92^{*}	0.81^{*}	0.95^{*}	0.90^{*}
	(0.37)	(0.39)	(0.38)	(0.39)
Distance			0.01	0.04^{**}
			(0.01)	(0.01)
Houston		0.01		-0.20
		(0.41)		(0.44)
King County		0.27		-0.12
		(0.55)		(0.58)
Los Angeles		1.27^{**}		1.16^{**}
		(0.40)		(0.42)
Orlando		0.59		0.52
		(0.45)		(0.48)
San Antonio		1.79^{**}		1.87^{**}
		(0.66)		(0.72)
San Jose		0.20		0.17
		(0.51)		(0.53)
Seattle		1.25^{**}		1.27^{**}
		(0.45)		(0.47)
Tucson		1.62^{***}		1.65^{***}
		(0.46)		(0.48)
Intercept	0.08	-0.80^{*}	-0.02	-1.00^{*}
	(0.13)	(0.40)	(0.15)	(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 2: *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.

correlation between race and jurisdiction does not drive the observed relationship. This pattern—while not necessarily causal—is precisely 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) and Cohen (2020).[7] 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 to measure such strength is to benchmark any potential unmeasured confounder against measured covariates in the model. In analysis presented below and 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 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 (2020) 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 would 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.

[7] Although the Cinelli and Hazlett (2020) 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.

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.

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

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 logistic 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 and a continuous variable of distance to closest trauma center in miles. Using this model we estimate the regression coefficient on White to be 0.70 (see the Appendix full regression specification results in Table 10), the associated fatality difference between White civilians and Black civilians controlling for city fixed effects. The lower bound estimate p follows from the estimated risk ratios (Cohen, 2020) as in equation

6 (see the appendix for the derivation).

$$p \geq 1 - \frac{1}{\hat{RR}} \tag{6}$$

Thus, we estimate 31% is the lower bound on the proportion of Black civilians that police would not have shot had they been White. Potentially there are unmeasured confounders that would affect both race and the likelihood of being fatally shot. As a sensitivity analysis we use both the binary indicator for Houston and the continuous variable of distance to closest trauma center as benchmarks. We estimate a lower bound adjusted for a possible unmeasured confounder of the same strength of association with race and with being fatally shot. Neither adjustment significantly decreases the estimated lower bound.

			Benchmark			
			Houston		**Distance**	
Approach	p	CI	p	CI	p	CI
Logistic (RR approach)	0.31	[0.21,1]	0.31	[0.12,1]	0.31	[0.2,1]

Table 3: *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. The Logistic RR approach uses the risk ratio to calculate and adjust the lower bound.

Substantively, our estimate of 31.3% is considerable and given it is a lower bound, may be much higher. Our estimate implies that police would not have shot 156 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, 2020).

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 aggression 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.[8] If Black civilians are aware (or believe) 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 police 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.

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.

As we noted above, we have not investigated Implication 2. Doing so would require

[8] Of course, Knox, Lowe, and Mummolo (2020) also suggest that the analysis in Fryer Jr. (2016) is flawed due to selection bias.

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. It 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. While doing so would require the collection of rich new data that are not currently available, we believe it is a worthy endeavor as scholars continue to work out the mechanisms underlying disparate outcomes in civilian-officer interactions.

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 use of compliance holds, pepper-spray, and canines) 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, 2020) 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 which civilians they interact with, 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 both the use of force continuum (e.g., no guns, guns drawn, guns fired) and outcomes (i.e., 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. Decentralization of law enforcement and varied discretion across localities in the United States further complicates research. Nonetheless, police departments, elected officials, and institutions of civilian oversight of police departments may become more interested in research about policing practices and outcomes, more anticipatory of scholarly needs, more transparent about and willing to share data with scholars and others through digitization and open-access, and interested in replication and extension of academic studies. If so, causal research on police behavior, from the spectacular to the mundane, may flourish, perhaps improving policymaking for public safety and improving policing (and police legitimacy) in the United States.

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.

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.

Bittner, Egon (1970). *The functions of the police in modern society: A review of background factors, current practices, and possible role models*. 2059. National Institute of Mental Health, Center for Studies of Crime and Delinquency.

Brown, Michael K. (1981). *Working the street: Police discretion and the dilemmas of reform*. Russell Sage Foundation.

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.

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.

Cohen, Elisha (2020). "Sensitivity Analysis on Outcome Test Lower Bound with Binary Data".

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 (2020). "Administrative Records Mask Racially Biased Policing". In: *American Political Science Review*, pp. 1–19. DOI: 10.1017/S0003055420000039.

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.

Lipsky, Michael (1980). *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. Russell Sage Foundation.

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.

Nieuwenhuys, Arne, Geert J.P. Savelsbergh, and Raôul 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.

Welch, Kelly (2007). “Black Criminal Stereotypes and Racial Profiling”. In: *Journal of Contemporary Criminal Justice* 23.3, pp. 276–288.

Wilson, James Q. (1978). *Varieties of Police Behavior*. Harvard University Press.

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 the 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, 2020).

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 rearrange 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 $\frac{\partial \pi(\tau)^*}{\partial k_r ho} > 0$, 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, 2020).

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:

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

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, 2020).

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:

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

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 subgame perfect Nash 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.[9]

[9]Empirical 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

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., $k_\rho < k_{\neg\rho}$), 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.

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

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.

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 subgame perfect Nash 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:

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

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

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

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 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 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:

$$\mathcal{F}(\rho) = \int\limits_{K(\rho)} (\delta(1)\pi^*(\tau) + \delta(0)(1-\pi^*(\tau))) \frac{\sigma^*(\tau)\, g(\kappa|\rho)}{\int\limits_{K(\rho)} \sigma^*(z|\rho) g(z|\rho) dz} d\kappa$$

$$\mathcal{F}(\rho) = \delta(0) + (\delta(1)-\delta(0)) \left(\frac{w_O + \delta(0)k_\rho}{1 + w_O + d_O - (\delta(1)-\delta(0))\, k_\rho} \right)$$

This implies that

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

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

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

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

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. □

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 subgame perfect Nash 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 $EU_i\left[s = 1, \pi|\tau\right] \geq EU_i\left[s = 0, \pi|\tau\right]$ and a sufficient condition for the officer to play $l = 1$ is $EU_O\left[l = 1, \sigma|\tau, s\right] \geq EU_O\left[l = 0, \sigma|\tau, s\right]$ Notice that if $s = 0$, the game ends. Therefore, we it is sufficient to show

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

which can be true for an arbitrarily large value of $c_O\left(\tau\right)$. 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

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

which can be true for an arbitrarily small value of $w\left(\tau\right)$. Now to see that the equilibrium is unique, notice that Assumption 2 and Lemma 1 imply the mixed strategies $\pi^*\left(\tau\right)$ and $\sigma^*\left(\tau\right)$ 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. □

C Data Cleaning and Organization

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 4a 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 must assume each row of data represents unique people. The incident in Table 4a 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 4b 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.[10]

Table 4: 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

[10] The 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, and our results do not change appreciably.

D 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 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:

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

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:

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

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:

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

D.1 Estimating the lower bound

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 logistic model:

$$P[\hat{Y} = 1|X] = \frac{exp^{\hat{\beta}_w D + \mathbf{X}\hat{\gamma} + \hat{\epsilon}}}{1 + exp^{\hat{\beta}_w D + \mathbf{X}\hat{\gamma} + \hat{\epsilon}}} \tag{13}$$

In equation 13, $\hat{\beta}_w$ is the coefficient of interest on the binary treatment variable D. We estimate it with a subset of officer-involved shooting data containing only Black and White civilians, where our main covariate of interest is an indicator for race (White equal to 1, Black equal to 0). $\mathbf{X}$ is a vector of covariates which include city fixed effects and distance in miles to closest trauma center. From this we can estimate the risk ratio of White compared to Black as

$$\widehat{RR} = \frac{E[P(\hat{Y} = 1|D = 1, \mathbf{X})]}{E[P(\hat{Y} = 1|D = 0, \mathbf{X})]} \tag{14}$$

Using this model we estimate the regression coefficient on White to be 0.70 (see Appendix Table 10 for the full regression specification results), the log odds of the associated fatality difference between White civilians and Black civilians controlling for city fixed effects and distance to closest trauma center. Using risk ratios we can estimate p (equation 15, see Cohen (2020) for the full derivation) the lower bound on the proportion of Black civilians that would not have been shot had they been white.

$$p \geq 1 - \frac{1}{\hat{RR}} \tag{15}$$

As an additional robustness check we also estimate a linear probability model (LPM). The results from the LPM can be used to estimate risk ratios as with the logistic or the coefficient on White can be used to directly estimate the lower bound (equation 16). The coefficient estimate on White is the numerator and the coefficient plus the observed mean fatality rate among Black civilians is the denominator. Using the logistic model we estimate 31% (31% and 32% for the LPM, see Table 5) as the lower bound on the proportion of Black civilians that police would not have shot had they been White.

$$p \geq \frac{\hat{\beta}_w}{F_b + \hat{\beta}_w} \tag{16}$$

			Benchmark			
			Houston		**Distance**	
Approach	p	CI	p	CI	p	CI
Logistic (RR approach)	0.31	[0.21,1]	0.31	[0.12,1]	0.31	[0.2,1]
LPM (RR approach)	0.31	[0.21,1]	0.30	[0.12,1]	0.31	[0.2,1]
LPM (partial R^2 adjustment)	0.32	[0.22,1]	0.29	[0.16,1]	0.32	[0.21,1]

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. The Logistic and LPM RR approach use the risk ratio to calculate and adjust the lower bound. The LPM partial R^2 adjustment uses Equation 7 to estimate p and uses the partial R^2 parameterization (Cinelli and Hazlett, 2020) to adjust the coefficient $\hat{\beta}_w$ by the benchmarks.

Substantively, our estimate of 31% is considerable and given it is a lower bound, may be much higher. Our estimate implies that police would not have shot 154 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.

E 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.

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.*

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 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.

Dependent variable:	Washington, DC Pat-down	NYC 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. $^{***}p \leq .001$, $^{**}p \leq .01$, $^{*}p \leq .05$

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.

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

G Sensitivity Analysis

Using the specifications from first column of Table 10 and 11 we conducted a sensitivity analyses in two ways. For the logistic model and LPM we use an observed covariate as a benchmark and estimate risk ratios of the associated effect of the covariate with the outcome and with treatment. From these risk ratios we calculate a bias factor (VanderWeele and Ding, 2017), adjust the main risk ratio on White and estimate a lower bound. This lower bound is now an an estimate of the proportion of Black civilians who would not have been shot had they been white, adjusted for a possible unobserved confounder of the same strength of association and as the benchmark covariate. We calculate a second adjusted lower bound for the LPM using a 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

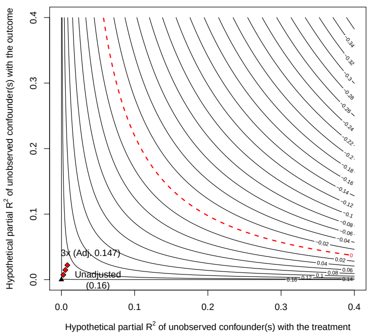
	Model 1	Model 2	Model 3	Model 4	Model 5
Black	-0.156^{***}	-0.200^{***}	-0.170^{***}	-0.157^{***}	-0.150^{***}
	(0.038)	(0.038)	(0.038)	(0.040)	(0.038)
Hispanic	0.019	0.063	0.015	0.021	0.026
	(0.039)	(0.038)	(0.039)	(0.040)	(0.039)
Asian/AI/AN/PI	0.174^{*}	0.225^{**}	0.178^{*}	0.169^{*}	0.192^{*}
	(0.080)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

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

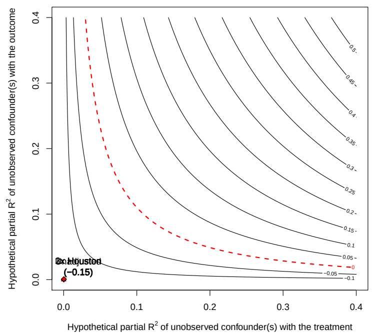
Table 9: Estimated relationship between civilian race and probability of fatality conditional upon being involved in an officer-involved shooting. Cells show linear model coefficients with standard errors. Omitted category is White civilians, Charlotte and the year 2010.

confounder against measured covariates in the model. Table 5 shows the results of the benchmark adjusted lower bounds and bootstrapped confidence intervals. This sensitivity analysis suggests that even an unobserved confounder of the same strength of the benchmark would not meaningfully change the estimate for the lower bound.

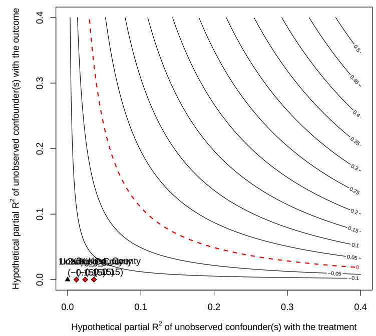
Each contour plot uses a different benchmark variable to show that any such unmeasured confounding variable would have to be more than three times as strong as any of variables we currently have in the LPM model (jurisdiction fixed effects, time fixed effects, and distance to trauma center). The red 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.



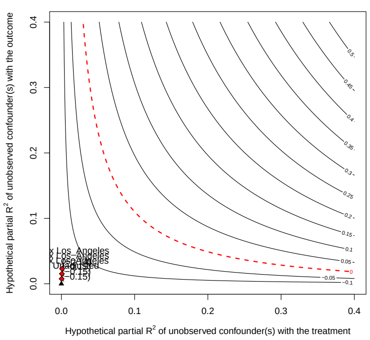
(a) Benchmark: Distance



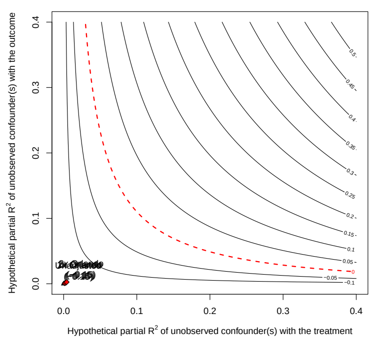
(b) Benchmark: Houston



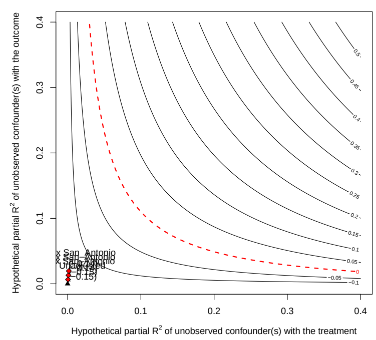
(c) Benchmark: King County



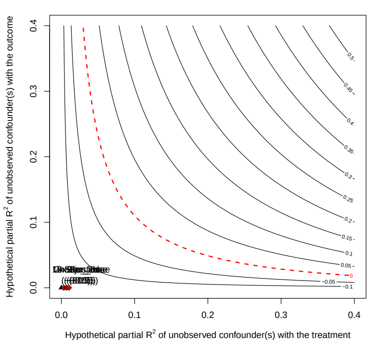
(d) Benchmark: Los Angeles



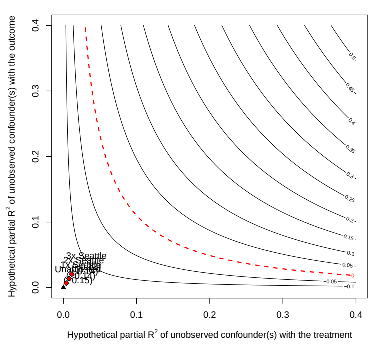
(e) Benchmark: Orlando



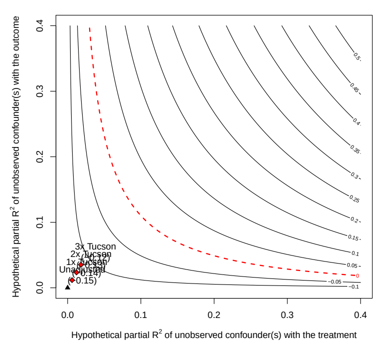
(f) Benchmark: San Antonio



(g) Benchmark: San Jose

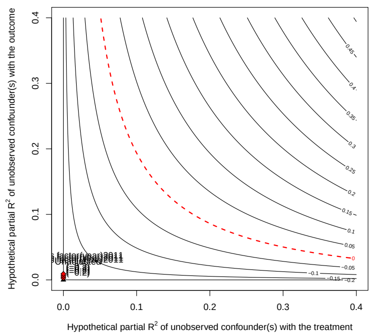


(h) Benchmark: Seattle

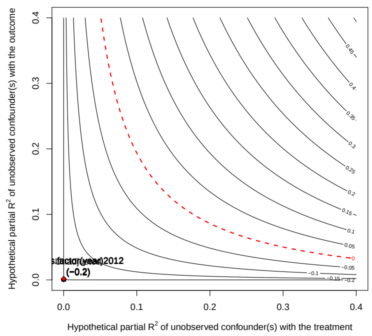


(i) Benchmark: Tucson

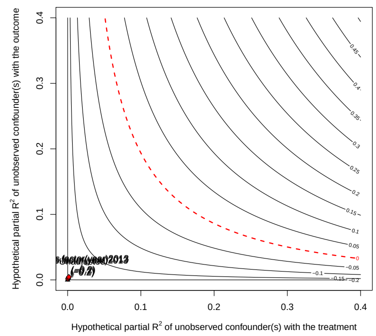
Figure 4: Sensitivity analysis using distance to trauma center and jurisdiction fixed effects as benchmarks



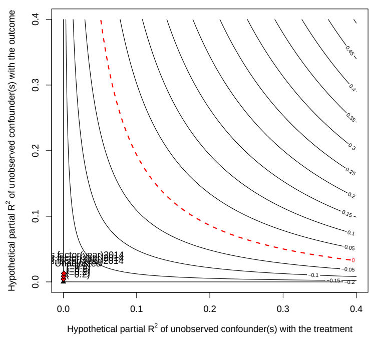
(a) Benchmark: Year 2011



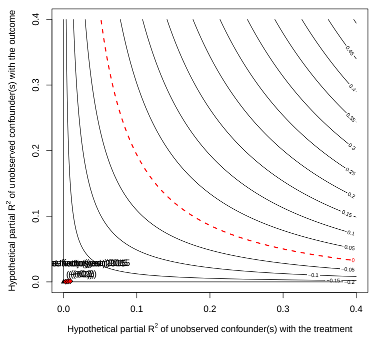
(b) Benchmark: Year 2012



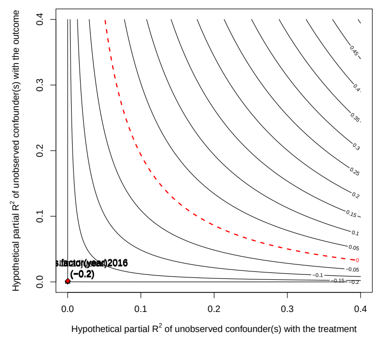
(c) Benchmark: Year 2013



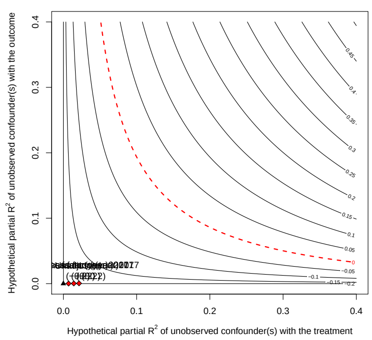
(d) Benchmark: Year 2014



(e) Benchmark: Year 2015



(f) Benchmark: Year 2016



(g) Benchmark: Year 2011

Figure 5: Sensitivity analysis using year fixed effects as benchmarks

DV:	Fatal	Fatal	Fatal
Data:	(Black/White)	(Hispanic/White)	(Asian/White)
White	0.70^{***}	-0.17	-0.80
	(0.17)	(0.19)	(0.41)
Charlotte	-1.57^{***}	-1.76^{*}	-0.97
	(0.42)	(0.81)	(0.90)
Houston	-1.65^{***}	-1.46^{***}	-0.93
	(0.24)	(0.28)	(0.54)
King County	-1.57^{**}	-0.94	-0.28
	(0.49)	(0.60)	(0.71)
Los Angeles	-0.45^{**}	0.04	0.69
	(0.17)	(0.16)	(0.43)
Orlando	-1.10^{***}	-1.53^{*}	-0.89
	(0.27)	(0.68)	(0.77)
San Antonio	14.08	0.49	14.99
	(429.74)	(0.61)	(623.80)
San Jose	-1.24^{*}	-0.74^{*}	0.06
	(0.53)	(0.36)	(0.70)
Seattle	-0.42	0.86	1.69^{***}
	(0.26)	(0.44)	(0.51)
Tucson	-0.55	1.22^{***}	1.53^{**}
	(0.33)	(0.34)	(0.56)
Closest Trauma (miles)	0.03^{*}	0.07^{***}	0.07^{*}
	(0.01)	(0.02)	(0.03)
N	719	746	281
AIC	918.32	950.15	351.27

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 10: Logistic 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.

DV:	Fatal	Fatal	Fatal
Data:	(Black/White)	(Hispanic/White)	(Asian/White)
White	0.16^{***}	-0.04	-0.16^{*}
	(0.04)	(0.04)	(0.08)
Charlotte	0.15	0.13	0.28
	(0.08)	(0.15)	(0.17)
Houston	0.13^{**}	0.17^{**}	0.29^{**}
	(0.05)	(0.06)	(0.11)
King County	0.14	0.29^{*}	0.43^{**}
	(0.10)	(0.13)	(0.15)
Los Angeles	0.40^{***}	0.52^{***}	0.65^{***}
	(0.04)	(0.04)	(0.09)
Orlando	0.25^{***}	0.17	0.30^{*}
	(0.06)	(0.13)	(0.14)
San Antonio	0.89^{***}	0.62^{***}	1.08^{**}
	(0.24)	(0.13)	(0.33)
San Jose	0.21	0.32^{***}	0.49^{**}
	(0.12)	(0.08)	(0.15)
Seattle	0.41^{***}	0.70^{***}	0.85^{***}
	(0.06)	(0.09)	(0.09)
Tucson	0.37^{***}	0.76^{***}	0.83^{***}
	(0.08)	(0.07)	(0.11)
Closest Trauma	0.01^{*}	0.02^{***}	0.01^{*}
Center (in miles)	(0.00)	(0.00)	(0.01)
Num. obs.	719	746	281
RMSE	0.47	0.47	0.46

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 11: 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.

Cohen, Elisha (2020). "Sensitivity Analysis on Outcome Test Lower Bound with Binary Data".

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 (2020). "Administrative Records Mask Racially Biased Policing". In: *American Political Science Review*, pp. 1–19. DOI: 10.1017/S0003055420000039.

VanderWeele, Tyler J and Peng Ding (2017). "Sensitivity analysis in observational research: introducing the E-value". In: *Annals of internal medicine*.

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.